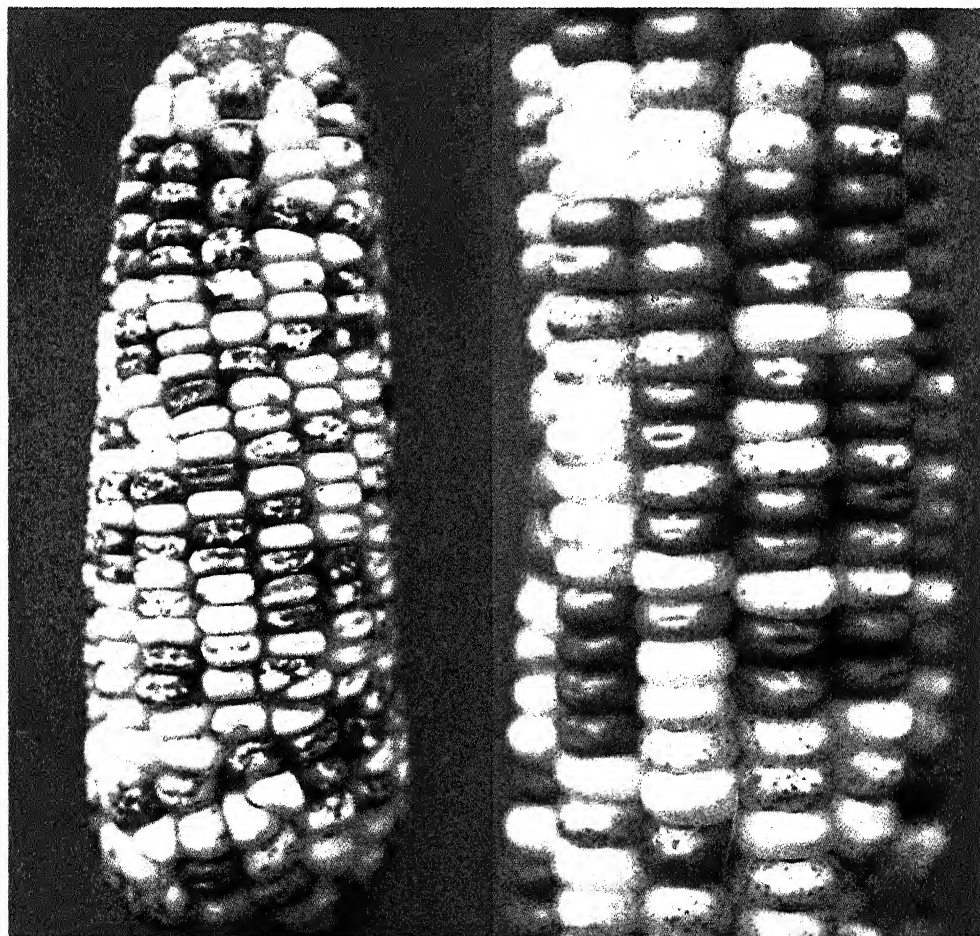


R e s o n a n c e

October 1996

Volume 1 Number 10

journal of science education



Barbara McClintock ❖ Database Mining ❖

Avogadro Number ❖ Fourier Series ❖ Reciprocal
Basis and Dual Vectors



**Resonance is a monthly journal of science education
published by Indian Academy of Sciences, Bangalore, India.**

Editors

N Mukunda (Chief Editor), *Centre for Theoretical Studies, Indian Institute of Science*
Vani Brahmachari, *Developmental Biology and Genetics Laboratory, Indian Institute of Science*
J Chandrasekhar, *Department of Organic Chemistry, Indian Institute of Science*
M Delampady, *Statistics and Mathematics Unit, Indian Statistical Institute*
R Gadagkar, *Centre for Ecological Sciences, Indian Institute of Science*
U Maitra, *Department of Organic Chemistry, Indian Institute of Science*
R Nityananda, *Raman Research Institute*
G Prathap, *Structures Division, National Aerospace Laboratories*
V Rajaraman, *Supercomputer Education and Research Centre, Indian Institute of Science*
A Sitaram, *Statistics and Mathematics Unit, Indian Statistical Institute.*

Corresponding Editors

S A Ahmad, Mumbai • H R Anand, Patiala • K S R Anjaneyulu, Mumbai • V Balakrishnan, Madras • M K Chandrashekar, Bangalore • Dhruvajyoti Chattopadhyay, Calcutta
• Kamal Datta, Delhi • S Dattagupta, New Delhi • S V Eswaran, New Delhi
• P Gautam, Madras • J Gowrishankar, Hyderabad • H Ila, Kanpur • J R Isaac, New Delhi
• J B Joshi, Mumbai • Kirti Joshi, Mumbai • R L Karandikar, New Delhi
• S Krishnaswamy, Madurai • Malay K Kundu, Calcutta • Partha P Majumder, Calcutta
• P S Moharir, Hyderabad • R N Mukherjee, Kanpur • M G Narasimhan, Bangalore
• S B Ogale, Pune • Mehboob Peeran, Bangalore • T P Radhakrishnan, Hyderabad
• G S Ranganath, Bangalore • Amitava Raychaudhury, Calcutta • P K Sen, Calcutta
• P N Shankar, Bangalore • Shailesh Shirali, Rishi Valley • V Srinivas, Mumbai
• R Srinivasan, Mysore • G Subramanian, Madras • V S Sunder, Madras
• R Tandon, Hyderabad • P S Thiagarajan, Madras • B Thimme Gowda, Mangalore
• R Vasudeva, Mysore • Milind Watve, Pune • C S Yogananda, Bangalore.

Assistant Editors Subashini Narasimhan, Sujatha Byravan

Production G Chandramohan **Editorial Staff** S Cecilia, G Madhavan,
T D Mahabaleswara, G V Narahari, M Srimathi **Circulation and Accounts** Peter Jayaraj,
Ranjini Mohan, B Sethumani, Shanthi Bhasker, B K Shivaramaiah, R Shyamala.

Editorial Office: Indian Academy of Sciences, C V Raman Avenue, PB No. 8005, Bangalore 560 080, India.

Tel: +91 (80) 3342310 / 3342546, Fax: +91 (80) 3346094, email: resonanc@ias.ernet.in

Editorial

N Mukunda, Chief Editor

Some three and a half centuries ago, Galileo Galilei declared that the book of nature was written in the language of mathematics. Isaac Newton following after him set up a comprehensive scheme for the description of physical phenomena, and aptly called it "Mathematical Principles of Natural Philosophy". For a couple of centuries thereafter physics and mathematics advanced largely in tandem, the needs of the one spurring ideas and advances in the other. In that era, some of the greatest contributors to mathematics were also outstanding physicists and conversely – to name but a few: Euler, Lagrange, Laplace, Gauss, Hamilton and Jacobi. Over the past century or so, however, there has been some parting of ways; one has come to acknowledge that each endeavour has its own character and spirit.

I refer to these matters now because in recent issues we have had articles both on mathematical topics *per se*, and on useful viewpoints and calculational techniques for the user of mathematics. Sometimes this has led to interesting discussions among us editors; there is a need to acknowledge that mathematics can be explained, understood and used at more than one level. No chance to recount an amusing anecdote should be lost, so here is one.

Physicists of an earlier generation would remember the names of Arthur S Wightman, Marvin L Goldberger and Geoffrey F Chew. It appears these three theorists were once discussing the way mathematics should be used in physics. Wightman felt that at any given point in time the best available resources and methods of mathematics must be used and the same demands of rigour should apply. Goldberger however differed, and said there should be room for physical intuition, the "sixth sense", to



"But it (the book of nature) cannot be read until we have learnt the language and become familiar with the characters in which it is written. It is written in mathematical language" —
Galileo Galilei



If not invented in the last century, the concept of matrices would have come into being through the laws of spectroscopy and Heisenberg's genius.

offset precision and rigour — "proof by persuasion" as he called it. And finally Chew asked "What is *proof* ?".

Apocryphal stories apart, one has to allow for these variations in style — chemists and physicists rely upon their "feel" of a problem and translate it into expectations of the mathematical expression of a solution. These are as real as anything else and cannot be denied or wished away. Here is a stunning instance from the history of modern physics. When Werner Heisenberg invented matrix mechanics in June 1925, he did not know what matrices were. Yet his deep physical intuition led him to represent physical position and momentum by arrays of complex numbers; and he then invented a law of multiplication for these arrays *based on the Ritz Combination Law of spectroscopy*. Later his teacher Max Born realized this was just the law of matrix multiplication, something *he* had learnt as a student. And in all honesty one can say — had the concept of matrices not been invented in the last century, it would have come into being through the laws of spectroscopy and Heisenberg's genius!

Two other striking examples come from Paul Dirac's work in quantum mechanics — his invention of the delta function, only much later accepted, formalised and christened as "distribution" in mathematics; and his rediscovery of spinors while constructing a relativistic wave equation for the electron.

Where then do we stand? Each of us must recall and admit what Hamlet said to his dear friend — "There are more things in heaven and earth, Horatio, Than are dreamt of in your philosophy". And ponder over this from Chen Ning Yang from more recent times: "It would be wrong, however, to think that the disciplines of mathematics and physics overlap very much; they do not. And they have their separate aims and tastes. They have distinctly different value judgements, and they have different traditions. At the fundamental conceptual level they amazingly share some concepts, but even there, the life force of each discipline runs along its own veins".



Science Smiles


R K Laxman



I have invented a wonder drug for the disease yet to be discovered, doctor!

SERIES ARTICLES

- 14**
- Translocation leading to Down Syndrome**
-
- The pedigree chart illustrates the inheritance of a reciprocal translocation between chromosomes 14 and 21. The first generation consists of a Grandfather (square) and a Grandmother (circle). The Grandfather is labeled '46 XY, Normal' and has two normal chromosomes 14 and 21. The Grandmother is labeled '46 X0, Homozygous Reciprocal Translocation' and has two chromosomes, each with a reciprocal translocation between 14 and 21. They have a daughter, the mother of the affected child, labeled '46 X0, Normal carrier'. She has one normal chromosome 14 and one chromosome with the reciprocal translocation. She and the Grandfather have a son, the affected child, labeled '46 X0, Down syndrome 7-day-old 21'. The affected child has two normal chromosomes 14 and 21. The pedigree also shows the parents of the mother: a father labeled '46 XY, Normal carrier' (square) and a mother labeled '46 X0, Normal carrier' (circle). Both parents have one normal chromosome 14 and one chromosome with the reciprocal translocation. They have a daughter, the mother of the affected child. At the bottom, the reciprocal translocation is shown as two chromosomes: one with a normal 14 and a translocated 21, and another with a translocated 14 and a normal 21.
- Grandfather
46 XY, Normal
- Grandmother
46 X0, Homozygous Reciprocal Translocation
- 46 XY, Normal carrier
- 46 X0, Normal carrier
- 46 X0, Down syndrome 7-day-old 21
- 46 XY, Normal carrier
- 46 X0, Normal carrier
- 14 21 14 21

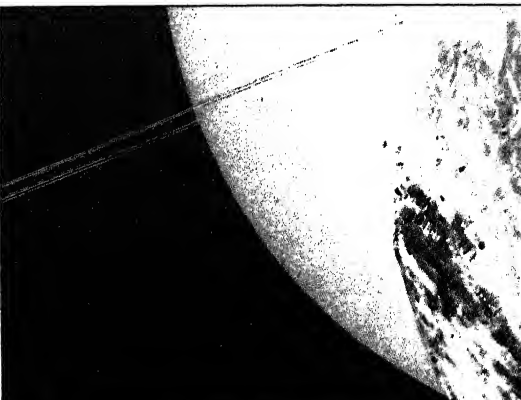
- 76
- 
- A black and white line drawing of a laboratory setup. In the center is a large microscope with a thick base, a curved arm, and a large eyepiece. A slide with a specimen is positioned on the stage. To the left of the microscope is a small bottle with a stopper. To the right is a beaker containing a liquid and a stirrer, and a small rectangular object, possibly a test tube or a small dish.

-

- 6** **What *Can* the Answer be?** *Reciprocal Basis and Dual Vectors* V Balakrishnan
- 14** **Know Your Chromosomes** *The Paths to Disorder Are Many* Vani Brahmachari
- 26** **Error Correcting Codes** *How Numbers Protect Themselves* Priti Shankar
- 37** **Learning Organic Chemistry Through Natural Products** *Architectural Designs in Molecular Constructions* N R Krishnaswamy

GENERAL ARTICLES

- 44 Fourier Series** *The Mathematics of Periodic Phenomena* S Thangavelu

- 26
- 

56 **Barbara McClintock and the Discovery of Jumping Genes** Vidyanand Nanjundiah

FEATURE ARTICLES

63 **What's New in Computers** *Database Mining* J R Haritsa

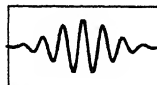
84 **RESEARCH NEWS**

- **How to Move in a Jostling Crowd** *The Art of Harnessing Random Motions* G S Ranganath

87 **BOOK REVIEWS**

- **Harmonic Analysis** *Fourier Series and Beyond*
K R Parthasarathy
- **Basic Biotechnology** S Vijaya

DEPARTMENTS



Editorial 1
Chief Editor's column/
ScienceSmiles *RKLaxman*



Classroom 76
Quantum Theory of the
Doppler Effect
G S Ranganath
Microbiology as if Bird
Watching *Milind G Watve*

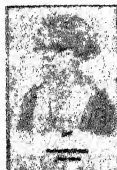


Think It Over 82
Counting Molecules in a
Spoonful of Water
J Chandrasekhar



Front Cover

Maize cobs displaying variegated Kernel phenotype. Variations are due to transposable elements. Photographs kindly provided by B M Prasanna of Indian Agricultural Research Institute, New Delhi.



Back Cover

Barbara McClintock (1902-1992)
(illustration by Prema)

What *Can* the Answer be?

2. Reciprocal Basis and Dual Vectors

V Balakrishnan



V Balakrishnan is a theoretical physicist engaged in teaching and research in the Department of Physics, Indian Institute of Technology, Chennai. His current research interests are in stochastic processes and chaotic dynamics. Among other frivolities, he has committed to memory large chunks from the works of Conan Doyle, Arthur Clarke and Wodehouse.

We usually express vectors as a sum of *basis vectors* which are mutually perpendicular and of unit length. In some situations, such as the description of crystals, it is necessary to use basis vectors which have any length and any angle between them. Solving for the coefficients in such an expansion introduces the concept of *reciprocal vectors* or *dual vectors*. They are the natural language to use in describing phenomena periodic in space, such as waves and crystal lattices. Generalisation of this concept to infinite dimensions leads to Dirac's notation for quantum states.

At the end of Part I of this series I stated that concepts such as reciprocal basis vectors, dual spaces and co-vectors could be motivated from simple considerations starting from well-known identities in elementary vector analysis.

Let us begin with the resolution of an ordinary vector \mathbf{v} in three-dimensional (Euclidean) space according to

$$\mathbf{v} = \mathbf{i} v_x + \mathbf{j} v_y + \mathbf{k} v_z . \quad (1)$$

What are v_x, v_y and v_z in terms of \mathbf{v} ? Clearly, $v_x = \mathbf{i} \cdot \mathbf{v}$, $v_y = \mathbf{j} \cdot \mathbf{v}$, $v_z = \mathbf{k} \cdot \mathbf{v}$. Therefore, if we introduce the *projection operator* $P_x = \mathbf{i}\mathbf{i}$ (no dot or cross in between the two vectors!), and 'operate' with it on the arbitrary vector \mathbf{v} by taking the dot product, the result $\mathbf{i}\mathbf{i} \cdot \mathbf{v}$ is defined to be precisely $\mathbf{i}(\mathbf{i} \cdot \mathbf{v}) = \mathbf{i} v_x$, the component or part of \mathbf{v} that lies along the unit vector \mathbf{i} . Similarly, we have projection operators $P_y = \mathbf{j}\mathbf{j}$ and $P_z = \mathbf{k}\mathbf{k}$. The *unit operator* (the operator that leaves any vector \mathbf{v} unchanged) is evidently just the sum of *all* the projection operators, namely,



$$\mathbf{I} = P_x + P_y + P_z = \mathbf{i}\mathbf{i} + \mathbf{j}\mathbf{j} + \mathbf{k}\mathbf{k}. \quad (2)$$

Thus Eq. (1) expresses the fact that

$$\mathbf{v} = \mathbf{I} \cdot \mathbf{v} = \mathbf{i}(\mathbf{i} \cdot \mathbf{v}) + \mathbf{j}(\mathbf{j} \cdot \mathbf{v}) + \mathbf{k}(\mathbf{k} \cdot \mathbf{v}). \quad (3)$$

We now ask the question: what is the counterpart of Eq. (3) in the case of *oblique axes* defined by three arbitrary, non-coplanar vectors \mathbf{a} , \mathbf{b} and \mathbf{c} (Figure 1), instead of the rectangular axes defined by \mathbf{i} , \mathbf{j} and \mathbf{k} ?

Once again, we can arrive at a solution by asking what the answer *can possibly be*. Writing

$$\mathbf{v} = \alpha \mathbf{a} + \beta \mathbf{b} + \gamma \mathbf{c}, \quad (4)$$

we observe that the coefficient α cannot involve any overlap¹ of \mathbf{v} with either \mathbf{b} or \mathbf{c} ; β cannot involve any overlap of \mathbf{v} with either \mathbf{c} or \mathbf{a} ; and γ cannot involve any overlap of \mathbf{v} with either \mathbf{a} or \mathbf{b} . Therefore α *must* be proportional to that part of \mathbf{v} which lies along $(\mathbf{b} \times \mathbf{c})$, i.e., to $[(\mathbf{b} \times \mathbf{c}) \cdot \mathbf{v}]$. Similar conclusions hold good for β and γ . Hence

$$\mathbf{v} = \lambda \mathbf{a} [(\mathbf{b} \times \mathbf{c}) \cdot \mathbf{v}] + \mu \mathbf{b} [(\mathbf{c} \times \mathbf{a}) \cdot \mathbf{v}] + \nu \mathbf{c} [(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{v}], \quad (5)$$

where the scalar factors λ , μ and ν are yet to be determined. The equivalence of all directions in space (the *isotropy of space*) suggests that λ , μ and ν must be equal to each other. This is easily borne out by setting $\mathbf{v} = \mathbf{a}$, \mathbf{b} and \mathbf{c} in turn. We find immediately that $\lambda = \mu = \nu = 1/[(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}]$. [Here we have used the cyclic symmetry of the scalar triple product, namely, $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} = (\mathbf{b} \times \mathbf{c}) \cdot \mathbf{a} = (\mathbf{c} \times \mathbf{a}) \cdot \mathbf{b}$.] Therefore

$$\mathbf{v} = \frac{\mathbf{a}[(\mathbf{b} \times \mathbf{c}) \cdot \mathbf{v}] + \mathbf{b}[(\mathbf{c} \times \mathbf{a}) \cdot \mathbf{v}] + \mathbf{c}[(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{v}]}{[(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}]} \quad (6)$$

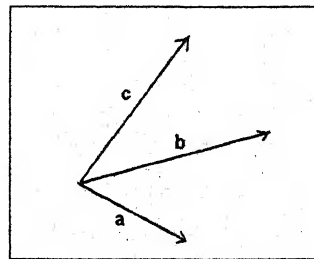


Figure 1 Oblique axes defined by a set of three arbitrary non-coplanar vectors \mathbf{a} , \mathbf{b} and \mathbf{c} .

¹ Overlap here means 'dot product' (also called 'scalar product' or 'inner product'). This step in the argument can be understood in more detail as follows. Let us keep α fixed and vary β and γ . The vector \mathbf{v} then varies over a plane parallel to the \mathbf{b} - \mathbf{c} plane. All vectors in this plane have the same value of α , but their projections on the \mathbf{b} - \mathbf{c} plane vary, so α cannot depend on those. It can only depend on the projection onto a vector normal to the \mathbf{b} - \mathbf{c} plane, that is, $\mathbf{b} \times \mathbf{c}$.

In three dimensional space, no more than three vectors of a given set of vectors can be linearly independent.

There is another, equally instructive, way to arrive at Eq. (6). We begin with the well-known vector identity

$$\mathbf{u} \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b}(\mathbf{u} \cdot \mathbf{c}) - \mathbf{c}(\mathbf{u} \cdot \mathbf{b}). \quad (7)$$

(A proof of Eq. (7) based on general arguments was given in Part I.) Now suppose \mathbf{u} is itself of the form $\mathbf{u} = \mathbf{v} \times \mathbf{a}$. Substitution in Eq. (7) gives

$$(\mathbf{v} \times \mathbf{a}) \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b}[(\mathbf{v} \times \mathbf{a}) \cdot \mathbf{c}] - \mathbf{c}[(\mathbf{v} \times \mathbf{a}) \cdot \mathbf{b}]. \quad (8)$$

The vector representing the quadruple cross product on the left-hand side is thus a linear combination of the vectors \mathbf{b} and \mathbf{c} . It therefore lies in the plane formed by these two vectors. However, we could as well have written $\mathbf{b} \times \mathbf{c} = \mathbf{d}$, in which case

$$\begin{aligned} (\mathbf{v} \times \mathbf{a}) \times (\mathbf{b} \times \mathbf{c}) &= (\mathbf{v} \times \mathbf{a}) \times \mathbf{d} \\ &= \mathbf{a}(\mathbf{v} \cdot \mathbf{d}) - \mathbf{v}(\mathbf{a} \cdot \mathbf{d}) \\ &= \mathbf{a}[\mathbf{v} \cdot (\mathbf{b} \times \mathbf{c})] - \mathbf{v}[\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})]. \quad (9) \end{aligned}$$

The *same* vector is therefore a linear combination of the two vectors \mathbf{a} and \mathbf{v} , and thus lies in the plane formed by them. As the four vectors \mathbf{v} , \mathbf{a} , \mathbf{b} and \mathbf{c} may be chosen quite arbitrarily, this appears to be paradoxical. However, we must now recall that these are vectors in three-dimensional space, *in which no more than three vectors of a given set of vectors can be linearly independent*, i.e., non-coplanar. In other words, if the vectors \mathbf{a} , \mathbf{b} and \mathbf{c} are a linearly independent set, the fourth vector \mathbf{v} *must* be expressible as a linear combination of these, precisely by equating the expressions found in Eqs. (8) and (9) and solving for \mathbf{v} . The result, after once again using the cyclic symmetry of the scalar triple product and some rearrangement, is precisely Eq. (6). This is the counterpart of the resolution in Eq. (3) of an arbitrary vector \mathbf{v} along orthogonal axes. The answer to our problem of resolving a vector \mathbf{v} in an arbitrary basis \mathbf{a} , \mathbf{b} , \mathbf{c} is thus

$$\mathbf{v} = \mathbf{a} (\mathbf{A} \cdot \mathbf{v}) + \mathbf{b} (\mathbf{B} \cdot \mathbf{v}) + \mathbf{c} (\mathbf{C} \cdot \mathbf{v}), \quad (10)$$

where

$$\mathbf{A} = \frac{\mathbf{b} \times \mathbf{c}}{V}, \quad \mathbf{B} = \frac{\mathbf{c} \times \mathbf{a}}{V}, \quad \mathbf{C} = \frac{\mathbf{a} \times \mathbf{b}}{V}, \quad (11)$$

with $V = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}$. The notation V arises from the fact that the modulus of $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}$ is the volume of the parallelepiped formed by the vectors \mathbf{a} , \mathbf{b} and \mathbf{c} . The vectors \mathbf{A} , \mathbf{B} and \mathbf{C} form the so-called *reciprocal basis*. The terminology is most familiar in crystallography: if \mathbf{a} , \mathbf{b} , \mathbf{c} are the primitive basis vectors of a lattice, \mathbf{A} , \mathbf{B} , \mathbf{C} are the basis vectors of the 'reciprocal' lattice. It is immediately verified that

$$\mathbf{A} \cdot \mathbf{a} = \mathbf{B} \cdot \mathbf{b} = \mathbf{C} \cdot \mathbf{c} = 1, \quad (12)$$

which helps explain why the term 'reciprocal basis' is used; also,

$$\mathbf{A} \cdot \mathbf{b} = \mathbf{A} \cdot \mathbf{c} = \mathbf{B} \cdot \mathbf{a} = \mathbf{B} \cdot \mathbf{c} = \mathbf{C} \cdot \mathbf{a} = \mathbf{C} \cdot \mathbf{b} = 0. \quad (13)$$

In fact, the reciprocal basis is *defined* in books on crystallography by Eqs. (12) and (13); the solutions are just the vectors in (11). It is easy to check that the general formula of Eq. (10) reduces to Eq. (3) in the special case of an orthogonal basis.

In what space do the reciprocal basis vectors (\mathbf{A} , \mathbf{B} , \mathbf{C}) 'live'? If the original basis vectors (\mathbf{a} , \mathbf{b} , \mathbf{c}) have the physical dimensions of length, Eqs. (11) show that (\mathbf{A} , \mathbf{B} , \mathbf{C}) have the physical dimensions of $(\text{length})^{-1}$. In crystallography and solid state physics this fact is used to define a 'reciprocal lattice' in wavenumber space, in which (\mathbf{A} , \mathbf{B} , \mathbf{C}) are the primitive lattice vectors. Why does one do this? It is not my intention to go into crystal physics here, but two good reasons (among several others) may be cited. In crystal physics, we have to deal very frequently with periodic functions, i.e., functions that satisfy $f(\mathbf{r}) = f(\mathbf{r} + \mathbf{R})$ where \mathbf{R} is any lattice vector $m\mathbf{a} + n\mathbf{b} + p\mathbf{c}$, and where m , n and p take on



integer values. Such a function can be expanded in a Fourier series of the form

$$f(\mathbf{r}) = \sum_{\mathbf{G}} f_{\mathbf{G}} \exp(i\mathbf{G} \cdot \mathbf{r}). \quad (14)$$

The summation over \mathbf{G} runs over *precisely the vectors of the reciprocal lattice*, i.e., $\mathbf{G} = h\mathbf{A} + k\mathbf{B} + l\mathbf{C}$, where (h, k, l) are integers. The second noteworthy point is that the Bragg condition for diffraction (of X-rays, electrons, neutrons, etc.) from a crystal is expressible in a very simple form in terms of \mathbf{G} , namely, $2\mathbf{k} \cdot \mathbf{G} = G^2$ (where \mathbf{k} is the wave vector of the incident beam). Likewise, the Laue conditions for diffraction maxima reduce to just $\mathbf{G} \cdot \mathbf{a} = h$, $\mathbf{G} \cdot \mathbf{b} = k$, $\mathbf{G} \cdot \mathbf{c} = l$ (which follow directly from Eqs. (12) and (13) and the definition of \mathbf{G}).

We are now at a point where the concepts of *ket* and *bra* vectors can be introduced naturally. Going back to Eq. (1), we note the following. Any vector \mathbf{v} can be represented in the form of a *column matrix* according to

$$\mathbf{v} = \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} v_x + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} v_y + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} v_z \quad (15)$$

(Here and in what follows, we shall freely use the '=' symbol between an abstract quantity and its *representation* in any form.)

To save space, let us write $(1 \ 0 \ 0)^T$ for the column matrix

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}. \text{ } T \text{ stands for 'transpose'. In this way of representing vectors,}$$

therefore,

$$\mathbf{i} = (1 \ 0 \ 0)^T, \mathbf{j} = (0 \ 1 \ 0)^T, \mathbf{k} = (0 \ 0 \ 1)^T. \quad (16)$$

We could also identify these with unit ket vectors denoted by $|e_1\rangle$, $|e_2\rangle$ and $|e_3\rangle$ respectively. Operating on a general vector $\mathbf{v} = (v_x, v_y, v_z)^T$, the projection operator $P_x = \mathbf{ii}$ introduced below Eq. (1) must yield the component $i v_x = (v_x \ 0 \ 0)^T$. This is achieved if we identify P_x with the 3×3 matrix $(1 \ 0 \ 0)^T (1 \ 0 \ 0)$. In other words, the i on the left in \mathbf{ii} really stands for the column vector $(1 \ 0 \ 0)^T$ or the ket vector $|e_1\rangle$, while the i on the right stands for the row vector $(1 \ 0 \ 0)$ – it is now natural to identify it with the bra vector $\langle e_1|$. The operator P_x is therefore $|e_1\rangle \langle e_1|$; similarly, $P_y = |e_2\rangle \langle e_2|$ and $P_z = |e_3\rangle \langle e_3|$. The ‘resolution of the identity’, Eq. (2), reads

$$|e_1\rangle \langle e_1| + |e_2\rangle \langle e_2| + |e_3\rangle \langle e_3| = \mathbf{I}. \quad (17)$$

The component v_x , which we saw was simply the scalar or dot product $\mathbf{i} \cdot \mathbf{v}$, is now written as the ‘inner product’ $\langle e_1| \langle \mathbf{v}$ where we have used the ket vector $|\mathbf{v}\rangle$ to denote the vector $\mathbf{v} = (v_x, v_y, v_z)^T$. We can then go on to generalize this idea of ket vectors and their “adjoint bra vectors” to n -dimensional Euclidean spaces, and then to infinite-dimensional Hilbert spaces. The whole treatment provides an admittedly heuristic, but easily digested, method of introducing the machinery of linear vector spaces (e.g., for quantum mechanics) to students of physics whose background in this regard comprises little more than some familiarity with elementary matrix analysis — the situation most commonly encountered.

Let us now translate our findings for oblique axes to this language of ket and bra vectors. Writing \mathbf{a} , \mathbf{b} and \mathbf{c} as the ket vectors $|a\rangle$, $|b\rangle$ and $|c\rangle$ respectively, Eq. (12) suggests at once that the reciprocal basis vectors \mathbf{A} , \mathbf{B} and \mathbf{C} are in fact to be identified with bra vectors $\langle A|$, $\langle B|$ and $\langle C|$. Equation (12) is the statement that the corresponding inner products are normalized to unity, i.e.,

$$\langle A|a\rangle = \langle B|b\rangle = \langle C|c\rangle = 1. \quad (18)$$

In crystallography, however, the structure of the lattice may force us to stick to the non-orthogonal basis as the natural and more useful one, supplemented by the reciprocal basis.

The expansion of an arbitrary vector \mathbf{v} in Eq. (10) reads, in this language,

$$|\mathbf{v}\rangle = (\langle \mathbf{A} | \mathbf{v} \rangle) |\mathbf{a}\rangle + (\langle \mathbf{B} | \mathbf{v} \rangle) |\mathbf{b}\rangle + (\langle \mathbf{C} | \mathbf{v} \rangle) |\mathbf{c}\rangle. \quad (19)$$

In other words, the resolution of the identity given by Eq. (17) for orthogonal coordinates is now replaced by

$$|\mathbf{a}\rangle \langle \mathbf{A}| + |\mathbf{b}\rangle \langle \mathbf{B}| + |\mathbf{c}\rangle \langle \mathbf{C}| = \mathbf{I}. \quad (20)$$

The space spanned by the reciprocal basis vectors \mathbf{A} , \mathbf{B} and \mathbf{C} (more generally, by bra vectors) may be regarded as a kind of *dual* of the original space spanned by the vectors \mathbf{a} , \mathbf{b} and \mathbf{c} . [This statement is a bit loose and glosses over certain technical details, but is quite acceptable at the present level of rigour.] It turns out that we can prove that the dual space is actually *isomorphic* to the original space, provided the latter is finite-dimensional (in our case, it is three dimensional). 'Isomorphic to' does not mean 'identical with', of course, but it does mean that the properties of the two spaces are essentially the same. This isomorphism between a linear vector space and its dual space *may* sometimes be valid even for infinite-dimensional spaces. A common but nontrivial example in physics occurs in elementary quantum mechanics: the position space wavefunction of a particle moving in one spatial dimension is a member of the linear vector space of square-integrable functions (of one real variable $x \in \mathbf{R}$). Its Fourier transform has a physical interpretation as the corresponding wavefunction in momentum space. This is also square-integrable, and is a member of an *isomorphic* vector space of square-integrable functions (of one real variable, $p \in \mathbf{R}$).

We have seen how 'reciprocal' vectors (in a 'dual' vector space) arise naturally if we work with an oblique set of axes. The distinction between the original space and the dual space exists in any case, but it may be blurred in the case of an orthogonal basis set like (\mathbf{i} , \mathbf{j} , \mathbf{k}) in a real vector space because the reciprocal



basis appears to coincide with the original basis. When faced with a non-orthogonal basis set, the usual practice in quantum mechanics is to construct an orthogonal basis by, say, the Gram-Schmidt procedure. In crystallography, however, the structure of the lattice may force us to stick to the non-orthogonal basis as the natural and more useful one, supplemented, as we have seen, by the reciprocal basis. It must be remembered that we have been working in three-dimensional Euclidean space for the greater part. What if the number of dimensions we have to deal with is not equal to three? (For one thing, the 'cross product' of two vectors is a vector only in three dimensions!) What if the space itself is curved? Do vectors and reciprocal vectors (or *bra* vectors), living in the dual vector space, have anything to do with the distinction between *contravariant* and *covariant* vectors, (or 'upstairs' and 'downstairs' indices), *tangent* and *cotangent* spaces, and maybe even the Lagrangian and Hamiltonian formalisms in classical mechanics? The answer is 'yes', implying that some profound aspects of the physical world are lurking behind the simple geometrical questions we have been discussing. We shall touch upon these matters in the next part of this series.

Some profound aspects of the physical world are lurking behind the simple geometrical questions we have been discussing.

Address for Correspondence
V Balakrishnan
Indian Institute of Technology
Chennai 600 036, India



100	99	98	97	96	95	94	93	92	91
65	64	63	62	61	60	59	58	57	90
66	37	36	35	34	33	32	31	56	89
67	38	17	16	15	14	13	30	55	88
68	39	18	5	4	3	12	29	54	87
69	40	19	6	1	2	11	28	53	86
70	41	20	7	8	9	10	27	52	85
71	42	21	22	23	24	25	26	51	84
72	43	44	45	46	47	48	49	50	83
73	74	75	76	77	78	79	80	81	82

Prime numbers occur along the diagonals in a square spiral arrangement of consecutive integers — Discovered by Stanislaw Ulam in 1963.

Know Your Chromosomes

4. The Paths to Disorder are Many

Vani Brahmachari



Vani Brahmachari is at the Developmental Biology and Genetics Laboratory at Indian Institute of Science. She is interested in understanding factors other than DNA sequence *per se*, that seem to influence genetic inheritance. She utilizes human genetic disorders and genetically weird insect systems to understand this phenomenon.

Mutations are vital in deciphering the regulated expression of genes in all organisms. In humans, study of mutant traits has helped in the management of genetic disorders. This article focuses on the nature of mutations that are encountered in general and the possible ways in which these mutations are produced.

From the gene mapping strategies discussed in the earlier articles of this series, you know that the trail of the gene is initially detected at the level of a disorder or an abnormality that runs in families. What I hope to do in this article is to give you a glimpse of the varied nature of the defects (mutations) that may ultimately result in the shutdown of an essential function of a cell, a tissue and therefore an organism and the probable ways in which these mutations are produced.

What Can Go Wrong With Genes?

The task at each cell division is that a 2 metre long DNA (or 6 billion units) has to be copied faithfully in every dividing cell in a limited time. The accuracy of the process is remarkable, otherwise genetic abnormalities and disorders would be more common than they are now. The reason behind this is that the system is able to handle errors, in a variety of ways. For example certain mutations being lethal lead to a reduction in the fertility of those carrying these mutations and the abnormal embryos get aborted in early stages of development.

The first step at which errors can arise as the fertilized egg begins to divide is a possible mispairing of bases during DNA

synthesis. This results in mismatches, with an adenine (A) pairing with some base other than a Thymine (T) and so on. The potential error rate can be as high as 1 to 10% per nucleotide, if we take into account the difference in free energy of pairs of complementary bases as against noncomplementary bases. But the observed frequency of mutation is much less, of the order of 1 in a million (10^{-6}) nucleotides. This high degree of fidelity of DNA replication is achieved by different components of the replication machinery, which can recognise and repair errors in replication.

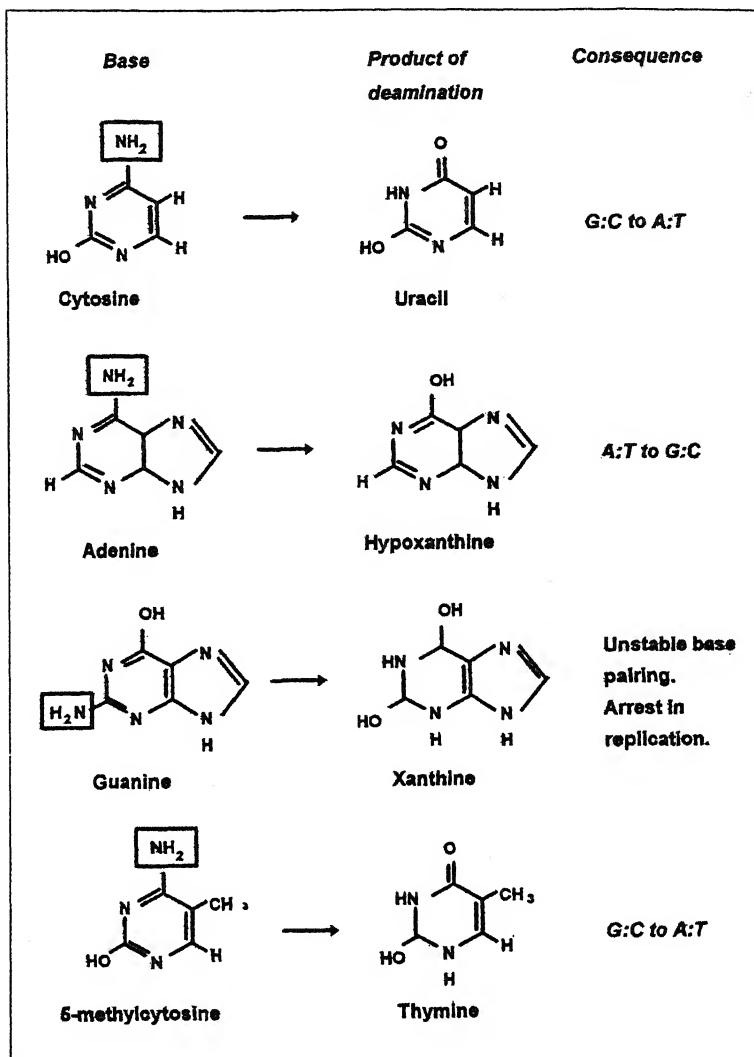
Certain mutations being lethal lead to a reduction in the fertility of those carrying these mutations and the abnormal embryos get aborted in early stages of development.

Another source of change in a base pair is the deamination of bases. For instance, if an amine group is removed from cytosine, it forms uracil, which can now pair with adenine during replication. Thus a C:G base pair is converted into a T:A base pair eventually, bringing about a mutation (*Figure 1*). Deamination of bases is known to occur spontaneously during DNA metabolism and is influenced by pH and temperature. The basic change is an alteration in one base pair, a C:G to T:A or vice-versa. A mutation of this sort is called a *point mutation*. The effect of a point mutation depends on where it occurs within the gene. It can be entirely without effect, in which case one may not even know that a change has taken place. Alternatively, a mutation can result in the protein coded by the gene becoming inactive or being unable to carry out its normal function.

There are instances where an additional length of DNA is inserted into the normal sequence. This may be due to the duplication of an existing sequence. Such an intrusion could scramble the normal message and therefore the protein that is made from that message. On the other hand insertion of a small number of bases into a coding region can lead to a different kind of disorder. If this number is neither three nor a multiple of three, the triplet code in the messenger RNA will be disturbed and the reading frame changed. This would result in the synthesis of protein with an aminoacid sequence different from



Figure 1 Products of deamination of bases commonly present in DNA and their consequences.



the wildtype protein, or the protein may terminate prematurely. This type of mutation is called a *Frame Shift* mutation. The removal of bases can also shift the codon reading frame. The reaction of free radicals with DNA can be a source of such errors (*Box 1*). There are several ways in which free radicals are generated in living systems. They interact with DNA and cause strand breakage and discontinuity.

All these types of mutations are seen in human genes. The probable mechanisms of mutations have been understood

RESONANCE

Questionnaire for Readers

1. Your age: < 15 15-20 20-25 25-30 30-40 40-50 50-60 >60

2. To which category do you belong?

- a) Student in high school XI/ XII/PUC/plus 2 BSc MSc
Engineering/Medicine PhD Other
- b) Teacher in VIII-X XI-XII/PUC/plus 2 UG PG Other
- c) Scientist in Univ./Research Institute, R & D/Industry Other
- d) Other (please specify)

3. Your subjects:

- | | |
|-------------------------|----------------|
| 1. Biology | 4. Mathematics |
| 2. Chemistry | 5. Physics |
| 3. Engineering/Medicine | 6. Other |

4. Do you subscribe to *Resonance*? Yes/No

5. How many issues of *Resonance* have you read so far?

1 2 3 4 5 6 7 8

6. Tell us what you read in *Resonance*

- | | |
|-----------------------------------|---------------|
| a) Biology | e) Physics |
| b) Chemistry | f) Other |
| c) Computer Science & Engineering | g) Everything |
| d) Mathematics | |

7. In what ways has *Resonance* helped you?

- | | |
|---------------------------------|------------------------------------|
| a) In understanding the subject | c) In increasing general knowledge |
| b) In your studies | d) Anything else (please specify) |

8. What do you think the length of an article (no. of pages) ought to be?

9. How would you rate the general quality of articles in each subject (in a scale of 1 to 5, 1 being poor, 5 being outstanding)? Circle your choice in each case

- | | | | |
|--------------------------|-----------|----------------|-----------|
| a) Biology | 1 2 3 4 5 | d) Mathematics | 1 2 3 4 5 |
| b) Chemistry | 1 2 3 4 5 | e) Physics | 1 2 3 4 5 |
| c) Computer Sci. & Engg. | 1 2 3 4 5 | f) Classroom | 1 2 3 4 5 |



10. In your opinion which categories of students are able to read and understand at least 50% of the articles?

- a) Plus two b) Undergraduate c) PG d) PhD

11. How much of the published material is directly usable in the classroom?

- <25% 25-50% >50%

12. What would you like to see more of in *Resonance*? (Tick as many as applicable)

- | | | |
|---------------------|------------------|----------------|
| a) Series articles | e) Experiments | i) Reflections |
| b) General articles | f) Book reviews | j) Any other |
| c) Features | g) Classroom | |
| d) Research News | h) Think it Over | |

13. At present the subscription to *Resonance* is highly subsidized. This may not be possible for a long period. How much do you think would be a fair subscription for 12 monthly issues of *Resonance*?

Individual, Rs.

Institutional, Rs.

14. Would you like *Resonance* to appear

- a) Monthly as now b) Once in two months

15. Any other comments:

(please be brief; use additional sheets if required.)

Name: _____

Address: _____

This form may please be completed and returned to:

The Chief Editor
Resonance
Indian Academy of Sciences
Post Box No. 8005
C V Raman Avenue
Bangalore 560 080, India



Box 1

Free Radicals and DNA

1. What are free radicals?

Free radicals are chemical species that contain one or more unpaired electrons, are capable of independent existence and are highly reactive. Example: hydroxyl free radical represented as OH^\cdot , superoxide radical ($\text{O}_2^{\cdot-}$).

2. How are free radicals produced in living organisms?

Exposure of organisms to ionizing radiation leads to fission of O-H bonds in water resulting in OH^\cdot . Oxidative stress due to several factors including cigarette smoking is believed to result in production of superoxide species, $\text{O}_2^{\cdot-}$.

3. Why are free radicals harmful?

Free radicals are highly reactive, they react with any molecule in their vicinity – proteins, lipids, carbohydrates and DNA. In DNA they lead to base modifications resulting in change of base pairs. They also react with sugar moieties which results in the deletion of a base and therefore leads to frameshift mutations. However not all free radicals are harmful. An oxide of nitrogen, nitric oxide NO^\cdot is a vasodilator and possibly an important neurotransmitter.

4. How do living systems handle free radicals?

Living systems have antioxidant defenses to remove $\text{O}_2^{\cdot-}$. Enzymes like superoxide dismutase convert $\text{O}_2^{\cdot-}$ into hydrogen peroxide (H_2O_2) and another enzyme, catalase converts H_2O_2 into water and molecular oxygen. It is interesting to note that a gene for superoxide dismutase is localized to chromosome 21, trisomy of which causes Down syndrome.

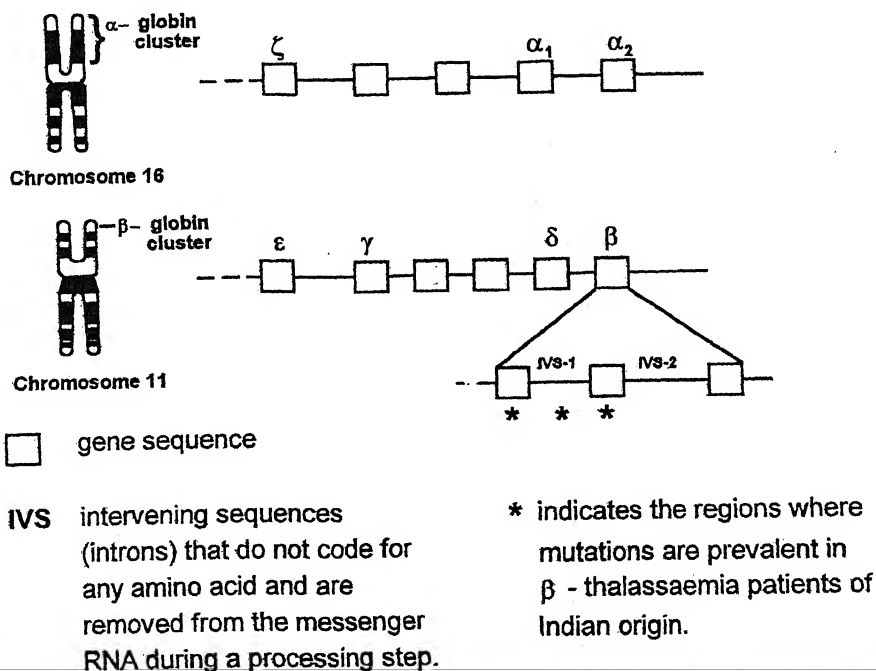
primarily by the study of bacteria, fungi and insects in which they can be induced by chemicals and radiation. One instance wherein different kinds of mutations occur and lead to similar disease states is haemoglobinopathies, the disorders due to defects in the protein globin in haemoglobin. The major groups of disorders are known as the thalassemias (α or β -thalassemia) (Box 2 and Table 1) — derived from the greek word *thalassa*, which means Mediterranean (from where many of these variations of globin gene originated). In humans, haemoglobin is made up of four protein chains of two different varieties α and β , and is represented as $\alpha_2\beta_2$. The gene coding for the α chain is on the short arm of chromosome 16, and that for the β chain is on the



Box 2

Mutations leading to Thalassaemias

Haemoglobin in humans is made up of four polypeptide chains designated as $\alpha_2\beta_2$. Mutations result in either β -thalassaemia or α -thalassaemia depending on whether the mutation is in the gene coding for β -globin chain or α -globin chain respectively. There are different types of haemoglobin at different stages of development, each adapted to oxygen requirement at these stages. The difference arises in the nature of polypeptide chains expressed from the α -globin gene cluster from chromosome-16 and β -globin gene cluster from chromosome-11. They differ in nucleotide sequence and are designated as ζ , α_1 , α_2 , ϵ , γ , δ , and β . There are several kinds of mutations that ultimately result in either reduced levels of haemoglobin or its total absence. Asian Indian, Chinese and African are the major ethnic groups at risk for thalassaemias.



short arm of chromosome 11. Defects in either of the two globin genes can result in reduced levels of haemoglobin or even in its total absence. Persons whose globin genes contain mutations of the kind described above are known to exist. The mutations affect the steps required to make a functional globin protein from the appropriate genes.

Table 1 Population studies have shown that within each ethnic group a certain mutation is more prevalent than others.

Type of mutation	Effect of mutation	Ethnic group where prevalent
Point mutation	Messenger not made	American black
	Defect in mRNA processing	Asiatic Indians & Chinese
	Incomplete protein made	Mediterraneans
Deletion	619 base pairs deleted, incomplete gene.	Asiatic Indians
	25 base pairs deleted; defect in mRNA processing	Data not available
Frameshift	Deletion of 1, 2 or 4 base pairs	Chinese

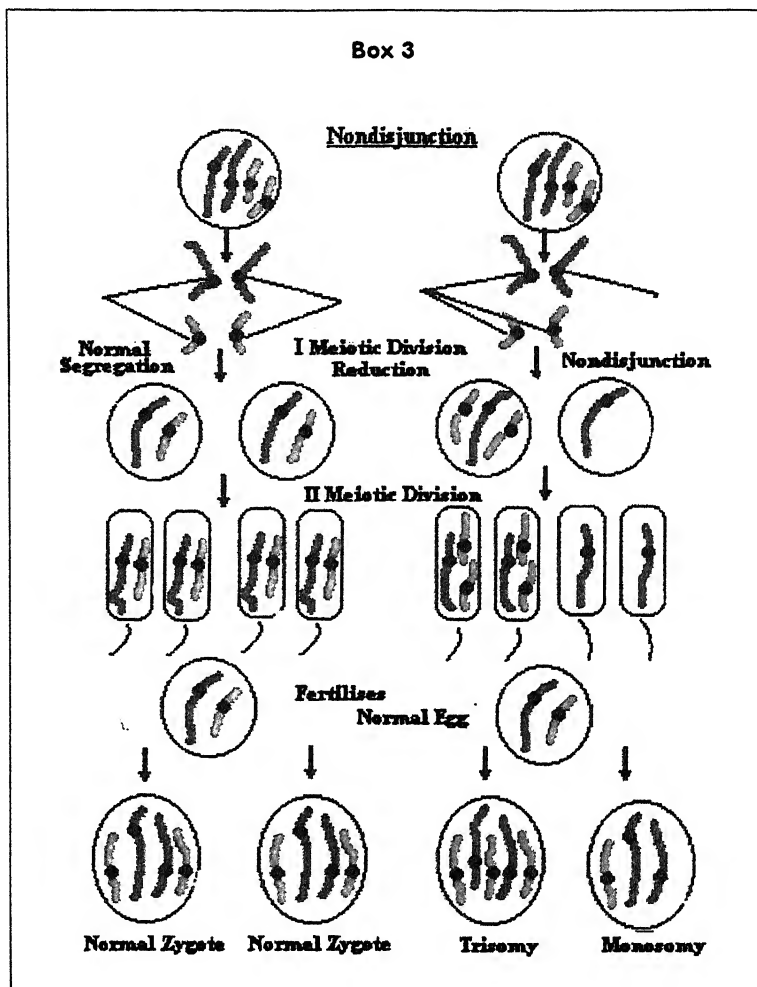
Data of Chakravarty, Purandare, Jaiswal and Gagati taken from the 7th International Conference on early prenatal diagnosis of genetic diseases (1994) and from *Essential Medical Genetics* by J M Connor and M A Ferguson-Smith.

Good Genes Needed But in the Right Number

The first chromosomal disorder to be recognised in humans was *Down's Syndrome*. This was earlier called *Mongolism*. Here chromosome 21, with its full complement of wild type or normal genes, is present in 3 copies instead of 2. Chromosome 21 belongs to the G group of chromosomes and is smaller than most other human chromosomes. The presence of abnormal chromosomal number described in general as aneuploidy, here trisomy, is observed in certain other syndromes too. Trisomies of chromosome 18, 13, 22, 8, 9 and X are known. Children with these 'numerical' anomalies have severe and complex malformations. Mental retardation is seen in all cases except in the trisomy of the X-chromosome. Anomalies in other chromo-



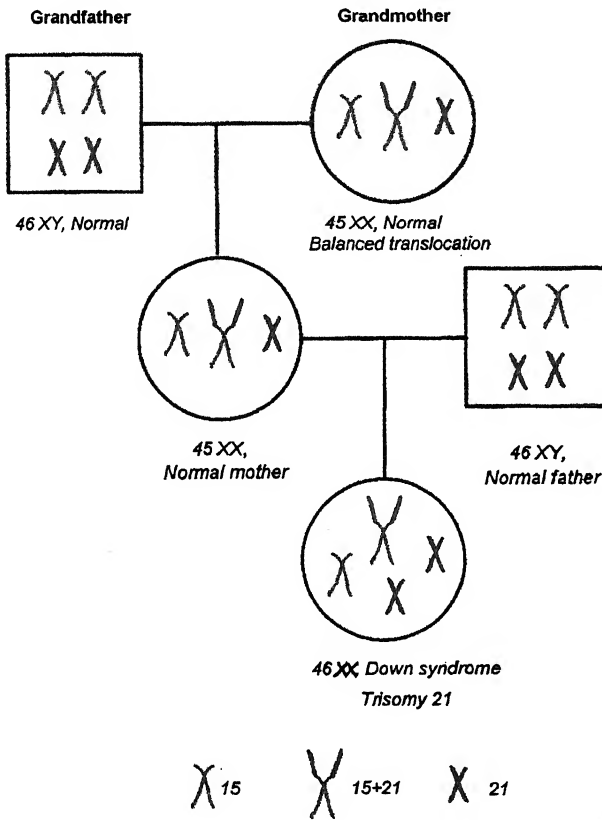
For simplicity diploid number is taken as two. Nondisjunction (ND) can result in trisomy or monosomy for a chromosome. The diagram here depicts nondisjunction at first meiotic division. Readers are encouraged to work out the consequence of ND at II meiotic division and ND for sex chromosomes at the I and II meiotic division.



somes are rarely seen in newborns, but are more frequent in natural abortions. This suggests that a deviation in the number of chromosomes disturbs normal embryonic development so much that the foetus is naturally aborted. In fact, a large fraction of natural abortions have chromosomal anomalies.

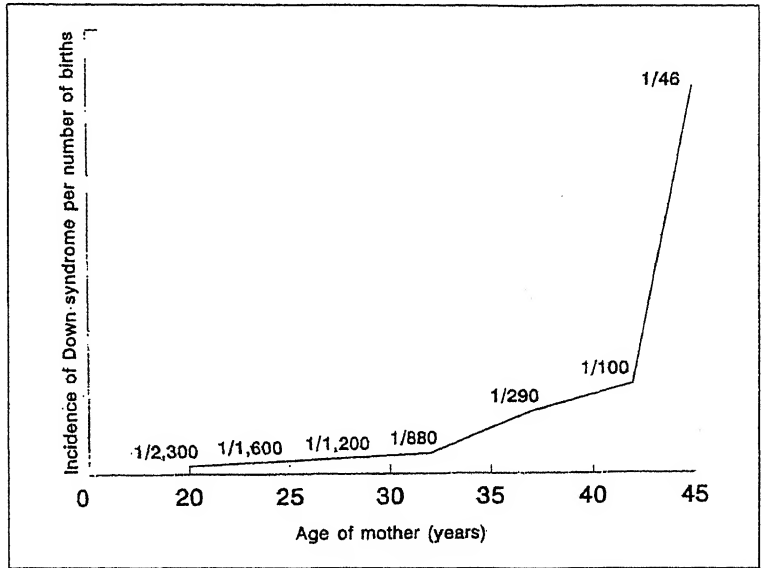
Similarly, a decrease in chromosome number from the diploid state can also lead to malformations. This is monosomy, there being only one chromosome instead of two. Monosomy for an X chromosome in the absence of a Y chromosome leads to female development but with malformations and mental retardation, a

Box 4 Translocation leading to Down Syndrome



situation known as *Turner syndrome*. The chromosome constitution of a Turner patient is 45 XO. The loss of the diploid state can also be caused by a deletion or removal of a gene or a part of the chromosome, in any one of two homologous chromosomes. Deletions of the short arm of the chromosome 18 and chromosome 5 are known to occur. The deletion of the short arm of chromosome 5 leads to the *Cri du Chat Syndrome* (cat cry syndrome). Children with this syndrome are not particularly malformed in their external features but have a striking cry resembling that of a cat and exhibit mental retardation.

Figure 2 The chances of birth of a Down syndrome child increase with age of the mother. (The figure is redrawn from *An Introduction to Genetic Analysis* by A J F Griffiths et al (1996) Sixth edition, W H Freeman and Company, New York).



Variations from the normal chromosome number can occur due to inappropriate separation during meiosis, which takes place before sperm and egg formation. Technically this is called *nondisjunction*. Similar segregation defects can lead to monosomy (Box 3). Sometimes chromosomes remain in the middle after metaphase and during anaphase (*Resonance* January 1996) and fail to segregate to the poles and enter the daughter nuclei. This is described as *anaphase lagging*. It is important to note that these are only terminologies that describe the phenomena but do not give any indication of mechanisms. We do not know how and why such disturbances in segregation are caused. In the case of trisomy 21 (Down syndrome) there is strong correlation between the age of the mother and the frequency of birth of a Down child (Figure 2), though there are cases where the extra chromosome 21 is contributed by the sperm. Therefore both an increased frequency of nondisjunction during meiosis in the mother and a decreased ability to reject abnormal embryos may contribute to the increased frequency of birth of Down children to older mothers. Thus perfectly normal parents can have normal children as well as a Down syndrome child. But when more than one Down child is born to young parents in a family, the clinician



would suspect reasons other than nondisjunction during egg or sperm formation. One such instance was described in 1960. The mother in this family had only 45 chromosomes but one of her two chromosome 15 was longer than the other. By banding, it was identified that a major part of chromosome 21 was attached to chromosome 15. She was normal as she had the correct complement of all chromosomes but one chromosome 21 was not free but attached to chromosome 15. The woman had inherited this from her mother. Her child had received one normal chromosome 21 from each of her parents along with the unusual chromosome (15 + 21) from her mother. Thus she had three copies of chromosome 21 instead of two and hence exhibited Down syndrome (*Box 4*). This brings us to yet another aberration in chromosomes namely *translocation* – basically meaning that parts of chromosomes change their location and move over to other chromosomes.

Variations from the normal chromosome number can occur due to inappropriate separation during meiosis, which takes place before sperm and egg formation. Technically this is called *nondisjunction*.

Abnormal Alliance of Chromosomal Regions Can be Unpleasant

Translocation of chromosomes can be harmless as in the mother and grandmother in the above example. These are described as *balanced translocation*. There are cases where two nonhomologous chromosomes exchange their parts but still maintain normal status with respect to their chromosome complement. This is described as a *balanced reciprocal translocation*. However when such chromosomes are passed on to the child along with a normal complement from the other parent it will result in aneuploidy (monosomy or trisomy) and lead to symptoms like developmental delay, malformations, mental retardation and congenital heart diseases.

These aberrations occur because of gaps and breaks within chromosomes which probably are caused by environmental factors like radiation and chemicals. The broken chromosomes can attach to their original counterparts but occasionally may join or



Mutations in germ cells will not affect the parent but can affect the children whereas somatic mutations can affect the individual but not the children.

get ligated to other nonhomologous chromosomes. Fusion of regions from different chromosomes may be brought about by recombination – like processes that take place during sperm and egg formation (*Resonance* March 1996).

The Philadelphia Chromosome

There are examples where a chromosomal translocation affects only certain cell types even though it is present in all the cells of the organism. In several patients with a type of leukemia called *chronic myelogenous leukemia* (CML), a balanced chromosomal translocation has been observed between chromosomes 9 and 22. This has apparently no effect in most of the tissues except in those of the circulatory system in which it induces leukemia. It was discovered by David Hungerford at the Fox Chase Institute in Philadelphia, U.S.A, based on careful cytology. The molecular basis for this has been traced to the activation of an oncogenic or cancer-inducing protein. Because of the translocation, 5 million base pairs of DNA originating from the end of chromosome 9 and carrying the cellular complement of an oncogene called *c-abl*, are translocated into a region of the long arm of chromosome 22. This results in expressing a protein which is a combination of the *c-abl* protein and the resident protein at this position on chromosome 22. The change activates the *c-abl* oncogene and leads to CML. This cancer is truly genetic, that is children of a CML patient have a finite chance (50%) of getting CML. But the question often asked is, are cancers genetic? Cancers are due to mutations at the DNA level, thus they are passed on from one cell to another, but this would not imply that the child of a cancer patient will also get cancer. For transmission from parent to child the mutation has to be in the germline of the parent, that is in the sperm or the egg cell (germline mutation). The other class of mutations which occur for instance in a liver cell of an individual leading to a malignant tumor is likely to be due to environmental abuses and is called *somatic mutation*. Therefore mutations in germ cells will not affect the parent but can affect

the children whereas somatic mutations can affect the individual but not the children.

Thus the right sequence of the gene, in the right dose and in its right neighbourhood are all essential for normal development and well being. Considering the number of ways in which the system can be derailed it is surprising that so many of us are considered 'normal'!

Note From the Author

The two previous two articles in the series had a list of genes on chromosomes 1 to 4. I plan to produce a poster containing the gene map of all human chromosomes later this year. Therefore the listing of genes and their functions is discontinued.

Acknowledgement

The author would like to thank Milind Kolatkar for all the illustrations.

Suggested Reading

- ◆ Alexander Macleod and Karol Sikora (ed). *Molecular Biology and Human Disease*. (Ed) Blackwell Scientific Publications, 1984.
- ◆ F Vogel and A G Motusky. *Human Genetics, Problems and Approaches*, II Edition. Springer-Verlag, Berlin, Heidelberg, New York and Tokyo, 1986.
- ◆ J M Connor and M A Ferguson-Smith. *Essential Medical Genetics – II* Edition. ELBS/ Blackwell Scientific Publications, 1987.
- ◆ EC Friedberg, G C Walker and W Siede. *DNA Repair and Mutagenesis*. ASM Press, Washington D. C, 1995.

Address for Correspondence
Vani Brahmachari
Developmental Biology and
Genetics Laboratory
Indian Institute of Science
Bangalore 560 012, India

Error Correcting Codes

1. How Numbers Protect Themselves

Priti Shankar



Priti Shankar is with the Department of Computer Science and Automation at the Indian Institute of Science, Bangalore. Her interests are in Theoretical Computer Science.

Linear algebraic codes are an elegant illustration of the power of Algebra. We introduce linear codes, and try to explain how the structure present in these codes permits easy implementation of encoding and decoding schemes.

Introduction

Many of us who have seen the pictures sent back by the spacecraft Voyager 2, have been struck by the remarkable clarity of the pictures of the giant gas planets and their moons. When Voyager sent back pictures of Jupiter's moons and its *Great Red Spot* in 1979, it was six hundred and forty million kilometres from Earth. A year later, it sent close up shots of Saturn's rings, clear enough to see the rotating spokes of the B-ring. In 1986 when transmitting near Uranus, Voyager was about 3 billion kilometres away, and in 1989, after being 12 years on the road to the outer reaches of the solar system, and nearly 5 billion kilometres away from Earth, it was able to transmit incredibly detailed, perfectly focused pictures of Triton, Neptune's largest moon. This great feat was in no small measure due to the fact that the sophisticated communication system on Voyager had an elaborate error correcting scheme built into it. At Jupiter and Saturn, a *convolutional* code was used to enhance the reliability of transmission, and at Uranus and Neptune, this was augmented by *concatenation* with a *Reed-Solomon* code. Each codeword of the Reed-Solomon code contained 223 bytes of data, (a byte consisting of 8 bits) and 32 bytes of redundancy. Without the error correction scheme, Voyager would not have been able to send the volume of data that it did, using a transmitting power of only 20 watts¹.

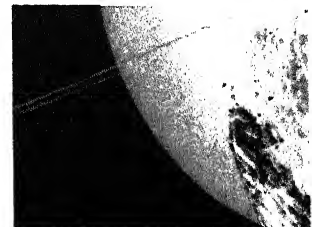
¹At the time of the writing of this article the Galileo space probe launched in 1989, has sent back stunningly sharp images of Ganymede, Jupiter's largest moon. The images, which came from Galileo's flyby of the moon on June 26–27 1996 are reported to be 20 times better than those obtained from the Voyager.

The application of error correcting codes in communication systems for deep space exploration is indeed impressive. However that is only part of the story. Error correcting codes are used extensively in storage systems. If you have a CD player at home, then you have a data storage system with error correction, that uses Reed-Solomon codes. A typical audio CD can hold up to seventy five minutes of music. The music is represented digitally, using zeros and ones. In fact, it takes about one and a half million bits to represent just one second of music. These bits are represented by pits on the mirror-like surface on one side of the disk. The pits have an average length of about one micron, and are recorded along a spiral track which is about 0.5 microns wide. Such microscopic features are susceptible to all kinds of errors - scratches, dust, fingerprints, and so on. A dust particle on the disk could obliterate hundreds of bits, and without error correction, cause a blast like a thunderclap during playback. However, around two billion bits are added, to protect the six billion or so bits on the disk. As a result, even if there is a 2mm long scratch(which corresponds to an error that spans around 2400 consecutive bits), the quality of music during playback will be *as good* as the original.

Error correcting codes are used in communication and storage systems.

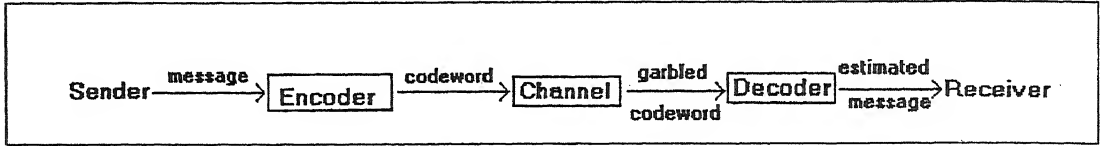
There is a conceptual kinship between the two examples here. Both involve the handling of data which may be corrupted by noise. There is a *source* of information, an *unreliable medium* or *channel* through which the information has to be transmitted, and a *receiver*. In the spacecraft, the medium is space, and the signals transmitted are subject to atmospheric disturbances, in addition to radiation. In the case of the CD, the communication is in *time* rather than in space, and the unreliable medium is the recording surface.

We owe to the genius of Claude Shannon, one of the finest scientific minds of this century, the remarkable discovery that reliable communication is possible, even over an unreliable medium. In 1948, Shannon, then a young mathematician, showed that arbitrarily reliable communication is possible at any rate



Reed-Solomon codes facilitated the transmission of pictures like this one from Voyager 2, a view of Uranus seen from one of its moons, Miranda.

Figure 1 Scheme for communicating reliably over an unreliable channel.



below something called the *channel capacity*. Roughly speaking, the channel capacity is its ultimate capability to transmit information, and one can push the channel to this limit and yet correct all the errors it makes using error correction. The illustration in *Figure 1* shows Shannon's scheme for communicating reliably over an unreliable channel. The message is sent over the channel, but before it is sent it is processed by an *encoder*. The encoder combines the message with some *redundancy*, in order to create a *codeword*. This makes it possible to detect or correct errors introduced by the channel. At the receiver's end, the codeword, which may be corrupted by errors is *decoded* to recover the message.

Let us assume that information is transmitted in terms of binary digits or bits. Given any sequence of k message bits, the encoder must have a rule by which it selects the r check bits. This is called the *encoding* problem. The $k+r$ bits constitute the codeword, and $n = k+r$ is the *block length* of the codeword. There are 2^n binary sequences of length n , but of these, only 2^k are codewords, as the r check bits are completely defined (using the encoding function), by the k message bits. The set of 2^k codewords is called the *code*. When one of these codewords is transmitted over the channel, it is possible that any one of the 2^n binary sequences is received at the other end, if the channel is sufficiently noisy. The *decoding* problem is then to decide which one of the 2^k possible codewords was actually sent.

Repetition Codes

Among the simplest examples of binary codes, are the *repetition codes*. These have $k=1$, r arbitrary, depending on the degree of



Claude Shannon showed that any communication process can be rendered reliable by adding protective redundancy to the information to be transmitted.

Claude Elwood Shannon was born in Gaylord, Michigan on April 30, 1916. His undergraduate work was in electrical engineering and mathematics at the University of Michigan, and he got his PhD in 1940 at the Massachusetts Institute of Technology. A year later he joined the Bell Telephone Laboratories in Princeton, New Jersey. Shannon's landmark paper 'A mathematical theory of communication' appeared in 1948. Essentially it said that it was possible to communicate perfectly reliably over any channel however noisy it may be. The result was breathtakingly original. Gradually, mathematicians and engineers began to realize that the problems of encoding, transmission, and decoding of information, could be approached in a systematic way. Thus was born a brand new field, which has since been greatly enriched by contributions from others.

error protection required, and $n = r + 1$. For example, if $r = 4$, we have a repetition code in which the symbol to be transmitted is repeated five times. There are two codewords, the all-zero codeword and the all-one codeword. At the receiving end, the decoder might use a majority vote of the bits to decide which codeword was actually transmitted. If there are more zeros than ones, then it decides that the all-zero codeword was sent, otherwise it decides that the all-one codeword was sent. If the channel noise flips more than half the bits in a codeword, then the decoder will commit a *decoding error*, that is, it will decode the received sequence into the wrong codeword. However, if less than half the bits are garbled during transmission, then the decoding is always correct. Thus, if $n = 5$, then two or fewer errors will always be corrected and the code is said to be *double error correcting*. It is easy to see that for arbitrary odd n , any combination of $(n-1)/2$ or fewer errors will be corrected. If the code has a long block length, and if channel errors occur infrequently, then the probability of a decoding error is very small. Let us compute the probability that a bit is decoded wrongly using a scheme where each bit is repeated five times.

Suppose we assume that the probability of a one being received as a zero or a zero being received as a one is p . This is sometimes called the *raw bit error probability*. We can compute the probability that a bit is decoded wrongly, using the decoding scheme where



the majority vote is taken. This is the probability that either three, four, or five bits are flipped by the channel. If we call this probability P_e , then this is given by

P_e = probability(3 errors) + probability(4 errors) + probability(5 errors)

Let N_i be the number of ways in which i bits can be chosen out of 5 bits.

$$P_e = N_3 p^3 (1-p)^2 + N_4 p^4 (1-p) + N_5 p^5$$

$$= 10p^3(1-p)^2 + 5p^4(1-p) + p^5.$$

If p is 0.1 we see that P_e is about 0.0086, giving a significant improvement. Smaller values for P_e can be obtained by increasing the number of repetitions. Thus we can make P_e as small as desired by increasing the length of the code. This is not particularly surprising, as all but one bit of each codeword are check bits. (Shannon's result promises much better).

The information rate R of a code is the ratio k/n and is a measure of the efficiency of the code in transmitting information. The repetition codes of block length n have rate $1/n$, and are the simplest examples of *linear algebraic codes*.

The Parity Check Matrix

Though the repetition codes have trivial encoding and decoding algorithms, they are not totally uninteresting. In fact, since they are so simple, they provide an easy illustration of a key concept in linear codes - the *parity check matrix* and the role that it plays in the decoding process.

We can consider each codeword to be a *vector* with n components in the *field* with the two elements 0 and 1 (We will refer to this field as F_2). The set of all vectors of length n over F_2 is a *vector space* with 2^n vectors. The code is a *subspace* with 2^k vectors. In the case of the repetition code of length five described

above, the vector space has thirty two vectors, that is, all combinations of five bits, and the code has two vectors.

In a linear algebraic binary code, the *sum* of two codewords is also a codeword. Addition of vectors is defined as *componentwise modulo 2* addition. In modulo 2 addition, $1+1=0$ and $0+1=1+0=1$. Thus the sum of vectors 10101 and 11001 is 01100. With modulo 2 addition, the sum and difference of two vectors is the same, as $+1$ is the same as -1 modulo 2. We write this as $1 \equiv -1 \pmod{2}$. The symbol \equiv is read as *is congruent to*. If two bits are the same then their sum is congruent to 0 modulo 2.

We follow a convention where the bits are indexed consecutively from left to right beginning with 0. Let $\mathbf{c} = (c_0, c_1, \dots, c_4)$ be a codeword for the length 5 repetition code.

The following four rules for a codeword completely specify the code.

1. The zeroth and first bits are the same.
2. The zeroth and second bits are the same.
3. The zeroth and third bits are the same.
4. The zeroth and fourth bits are the same.

The above rules are equivalent to the following four equations

$$c_0 + c_1 \equiv 0 \pmod{2}$$

$$c_0 + c_2 \equiv 0 \pmod{2}$$

$$c_0 + c_3 \equiv 0 \pmod{2}$$

$$c_0 + c_4 \equiv 0 \pmod{2}$$

The equations above can be expressed in matrix form as

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

In medical parlance, a syndrome is an indicator of underlying disease. Here too, a non zero syndrome is an indication that something has gone wrong during transmission.

The first matrix on the left hand side is called the *parity check* matrix H . Thus every codeword c satisfies the equation

$$Hc^T = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Therefore the code can be described completely by specifying its parity check matrix H .

Problem

What is the parity check matrix for a binary repetition code with two message bits, each of which is repeated three times?

Decoding

The decoder has to eventually answer the following question: Given the received pattern, what is the most likely transmitted codeword? However, it turns out that it is easier to focus on an intermediate question, namely: What is the estimated error pattern introduced by the noisy channel? For if r is the received pattern, and e is the error pattern, and addition between patterns is defined componentwise modulo 2, then $r = c + e$, and therefore, $c = r + e$. (In the binary case, $r + e = r - e$). If the receiver gets the pattern r and he forms the product Hr^T , then this is, $Hr^T = Hc^T + He^T = He^T$. In general, this is not the all zero column vector, though of course, if the error pattern corresponds to a codeword, it *will* be the all zero vector. This product is of key importance in decoding and is called the *syndrome*. Note that the syndrome bits reveal the pattern of parity check *failures* on the received codeword.

In medical parlance, a syndrome is an indicator of underlying disease. Here too, a non zero syndrome is an indication that something has gone wrong during transmission. We can carry the analogy even further. A syndrome does not point to an



unique disease. Here, the same syndrome can result from different error patterns. (Try the error patterns 11000 and 00111, for instance). The job of the decoder is to try to deduce from the syndrome which error pattern to report.

Suppose a single error has occurred during transmission. Then the error pattern will have a single 1 in the position in which the error has occurred and zeros everywhere else. Thus the product He^T will be the i^{th} column of H if i is the error position. Since the five columns of H are all distinct, each single error will have a distinct syndrome which will uniquely identify it. If the fifth bit is in error, the error pattern is 00001, and the syndrome will be the column of H with entries 0001.

What happens if there is a double error? Continuing the line of argument above, each of the two 1's in a double error pattern will identify an unique column of H and therefore the syndrome for a double error pattern, with 1's in positions say i and j , will just be the mod 2 sum of columns i and j of H . The number of double error patterns here is ten. (This is the number of ways in which 2 bits can be chosen from 5). The list of syndromes for the ten double error patterns is: 1110, 1101, 1011, 0111, 0011, 0101, 1001, 0110, 1010, 1100. That for the five single error patterns is: 1111, 0001, 0010, 0100, 1000. We can see that every single and double error pattern has a distinct syndrome that uniquely identifies it.

The syndromes for the single and double error patterns, together with the all zero syndrome, account for all the sixteen combinations of the four bits of the syndrome. Thus the syndrome for any error pattern with three or more 1's, will be the same as the syndrome for some error pattern with up to two 1's. For example, 00111 gives the same syndrome as 11000. How does the decoder decide which error pattern to report?

Under the assumption that the raw bit error probability is less than $1/2$, and that bit errors occur independently of one another,

A syndrome does not point to an unique disease. Here, the same syndrome can result from different error patterns.



the more probable error pattern is the one with fewer 1's. Thus the decoder follows what is known as a *maximum likelihood* strategy and decodes into the codeword that is *closer* to the received pattern. (In other words, it chooses the error pattern with fewer 1's). Therefore, if a table is maintained, storing the most likely error pattern for each of the sixteen syndromes, then decoding consists of computing the syndrome, looking up the table to find the estimated error pattern, and adding this to the received message to obtain the estimated codeword. For most practical codes, storing such a table is infeasible, as it is generally too large. Therefore estimating the most likely error pattern from the syndrome is a central problem in decoding.

Problem

How does syndrome decoding of the repetition code of length 5 compare in complexity with majority-vote decoding?

Hamming Geometry and Code Performance

The notion of a pattern being 'closer' to a given pattern than another may be formalized using the *Hamming distance* between two binary vectors. The Hamming distance between two vectors is the number of positions in which they differ. For example, the Hamming distance between the vectors 11001 and 00101 is three, as they differ in positions 0,1 and 2. The *Hamming weight* of a vector is the number of non-zero components of the vector. For example, the Hamming weight of the vector 10111 is four. The *minimum distance* of a code is the minimum of the Hamming distances between all pairs of codewords. For the repetition code there are only two codewords, and the minimum distance is five.

Problem

For each length n describe the code of largest possible rate with minimum distance at least two.
(Hint: Each codeword has one check bit, and the parity check matrix will have just one row).

The Hamming distance between two vectors is the number of positions in which they differ. The Hamming weight of a vector is the number of non-zero components of the vector.



We saw for the repetition code of length 5, that if a codeword is corrupted by an error pattern of Hamming weight 2 or less, then it will be correctly decoded. However, a weight of three or more in the error pattern results in a decoding error. In fact, if the weight of the error pattern is five, the error will go by undetected. This example hints at the following result:

If the Hamming distance between all pairs of codewords is at least d , then all patterns of $d-1$ or fewer errors can be detected. If the distance is at least $2t + 1$, then all patterns of t or fewer errors can be corrected.

Figure 2 illustrates this result for the repetition code of length 5. The minimum distance of this code is 5, and by the result above, it should be able to correct all errors of Hamming weight up to 2.

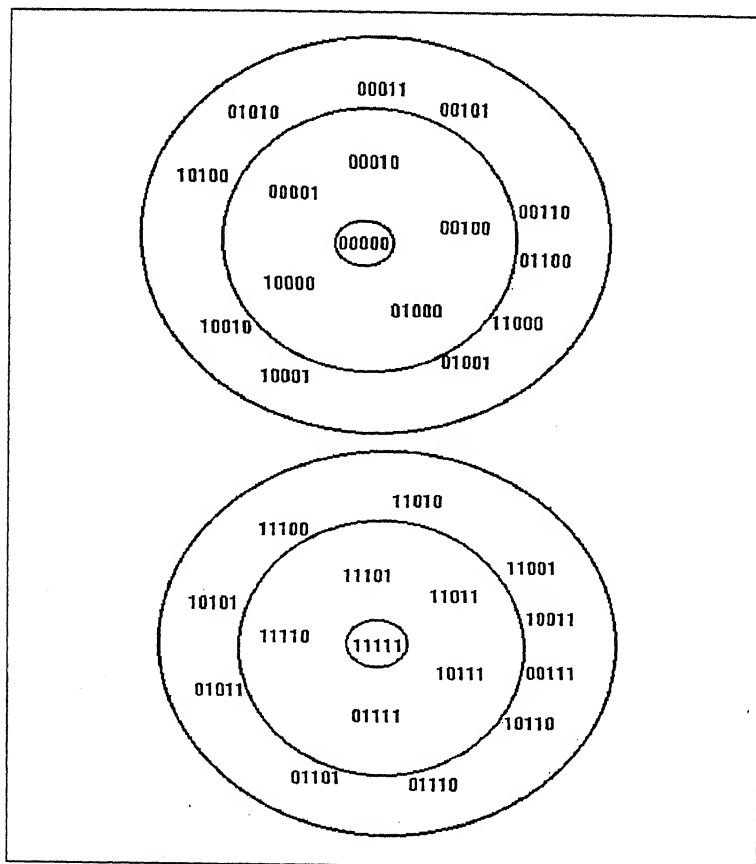


Figure 2 Hamming spheres of radius 1 and 2 around codewords for the binary repetition code of length 5.

Lemma-rick
by
S W Golomb

A message with content
and clarity
Has gotten to be quite a
rarity
To combat the terror
Of serious error
Use bits of appropriate
parity.

Thus if any codeword is altered by an error pattern of weight up to 2, then the resultant pattern should be distinct from one obtained by altering any *other* codeword with an error pattern of weight up to 2, (else the two error patterns would have the same syndrome, and would hence be indistinguishable). Taking a geometric view of the code, if spheres are drawn around all codewords, each sphere containing all vectors at distance 2 or less from the codeword, then the spheres will be non intersecting. In fact, the spheres cover the *whole* space, that is, they together contain *all* the vectors in the space. This generally does not happen for all codes. The repetition codes happen to belong to a very exclusive class of codes called *perfect codes*.

Problem

Show that the geometric property described above holds for any repetition code of odd length n , that is, the Hamming spheres of radius $(n-1)/2$ around codewords cover the whole space.

In the next article we will study the single error correcting Hamming codes invented in 1948. Apart from having a simple and elegant algebraic structure, these are historically important, as they were the earliest discovered linear codes. A great deal of work in constructive coding theory followed the appearance of Hamming's pioneering paper in 1950. But it was only ten years later that a general construction for double error correcting codes was discovered, and this was soon generalized to t -error correcting codes, for all t .

Suggested Reading

- ◆ **Robert J McEliece.** *The Theory of Information and Coding. Encyclopedia of Mathematics and its Applications.* Addison Wesley, 1977. An introduction to the subject of information theory and coding. Perhaps the most readable text in the area.
- ◆ **J H Van Lint.** *Introduction to Coding Theory (Second edition).* Graduate Texts in Mathematics. Springer-Verlag, 1992. A rigorous, mathematically oriented, compact book on coding, containing a few topics not found in older texts.

Address for Correspondence
Priti Shankar
Department of Computer
Science and Automation
Indian Institute of Science
Bangalore 560 012, India

Learning Organic Chemistry Through Natural Products

6. Architectural Designs in Molecular Constructions

N R Krishnaswamy

We will discuss in this article an outline of the biosynthesis and a few laboratory-designed syntheses of camphor.

In the first two articles of this series we saw how the structures and conformations of naturally occurring compounds could be determined. In the third and fourth we discussed the relationships between structure on the one hand and chemical and biological properties on the other. The fifth was designed to give you some practical experience in isolation of a few of these compounds from their natural sources. In this sixth and final part of the series we shall illustrate with the help of an example how the construction of a molecule is designed and executed in nature and in the laboratory.

Total laboratory synthesis and biosynthesis of natural products are complementary in the sense that if one is the end product of human intellect, ingenuity and technical and experimental skills, the other is the result of what one may sum up as the total wisdom of nature evolved over millions of years. We do not know how this wisdom was evolved and what the motivating forces were. We also do not know whether the chemistry of the secondary metabolites present in plants, insects, animals etc of the present time is the same as that of the earlier generations and whether it would remain the same in the future. We can only guess that the answer is 'perhaps not', if the understanding of the role of metabolites is correct. With the changes in the environment, however subtle and slow, there have to be corresponding changes



N R Krishnaswamy was initiated into the world of natural products by T R Seshadri at University of Delhi and has carried on the glorious tradition of his mentor. He has taught at Bangalore University, Calicut University and Sri Sathya Sai Institute of Higher Learning. Generations of students would vouch for the fact that he has the uncanny ability to present the chemistry of natural products logically and with feeling.

Box 1

For a small molecule, camphor, perhaps, is one of the most agile and articulate compounds known. It has a characteristic smell and is widely used as deodorant and insect repellent. Camphor is optically active and both the (+) and (-) forms as well as the racemate occur in nature. (+) Camphor is obtained from the wood of the camphor tree, *Cinnamum camphora* where-as the (-) form is isolated from the essential oil of *Matricaria parthenium*. Racemic camphor occurs in *Chrysanthemum sinensis*.

in the strategies used for self-defence and survival of the species. We do hope that some of our young readers would think of this and devise appropriate methods of investigation which could give answers to these questions. As was pointed out earlier (in different terms) present day chemistry of natural products is equivalent to lifting the curtains that block our view of the drama being continuously enacted in nature; only a part of the action has been unveiled. We see only what has been created and is existing. What is more interesting is to know who or what (if an answer to this question could ever be found!), why and how the script was prepared for this grand show. That is, indeed, a tall order and a challenge for the bravest and most daring!

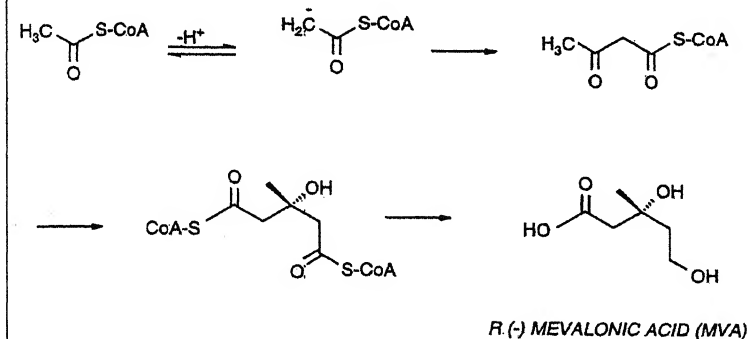
We will discuss in this article an outline of the biosynthesis (omitting biochemical details) and a few laboratory-designed syntheses of camphor. This important and well known naturally occurring ketone (*Box 1*) is a monoterpene and is biosynthetically derived from R (-) mevalonic acid. Mevalonic acid itself is synthesised from three acetic acid units. Each acetic acid is first 'activated' by conversion into acetyl coenzyme-A. In this step we see that nature is aware of the need to replace hydroxyl of the carboxylic acid with a better leaving group before the carbonyl of the carboxyl function can easily be attacked by a nucleophile. In the laboratory a carboxyl group is usually activated by conversion into its anhydride or acid chloride (*Box 2*). Nature uses coenzyme-A which has a sulfhydryl (SH) group which undergoes acetylation. Acetyl coenzyme-A can also lose a proton to generate a carbanion which then attacks the carbonyl group of another

Box 2

The acid chloride and the anhydride are more reactive than the free acid because the chloride and carboxylate ions are better leaving groups than OH⁻. Activation of the carboxyl group is an essential prerequisite in peptide synthesis.



Scheme-1



acetyl coenzyme-A molecule as shown in *Scheme-1*. The product is a thioester which also has a keto carbonyl group. Reaction of this compound with the carbanion of a third acetyl coenzyme molecule at the keto carbonyl group results in a compound having six carbon atoms. This reaction is a biochemical equivalent of aldol condensation (*Box 3*). Reduction by hydride ion transfer (*Box 4*) of one of the thioester groups to the corresponding alcohol and hydrolysis of the other leads to mevalonic acid as shown in the Scheme. The compound has an asymmetric carbon and therefore, two enantiomorphous forms are possible. Natural mevalonic acid has the R configuration and is laevo-rotatory (*Box 5*)

Scheme-2 shows the biochemical conversion of mevalonic acid into isopentenyl pyrophosphate and then on to geranyl pyrophosphate. This transformation takes place in several steps. In the first step, the primary hydroxyl of mevalonic acid is converted into its pyrophosphate. This reaction is brought about by an enzyme which requires adenosine triphosphate (ATP). Mevalonic acid pyrophosphate then undergoes dehydrative decarboxylation in the presence of ATP to yield isopentenyl pyrophosphate (IPP). Isomerization of IPP results in dimethylallyl pyrophosphate (DMAPP). Combination of one IPP molecule with one DMAPP then gives geranyl pyrophosphate.

Box 3

The aldol condensation is one of the widely used reactions in preparative organic chemistry. The classical example of this reaction, which can be catalysed either by a base or by an acid, is the self condensation of two molecules of acetaldehyde. A crossed aldol reaction involving two different aldehydes or ketones is also known as the Claisen-Schmidt reaction.

Box 4

Reductions by hydride ion transfer occur in nature through the mediation of the coenzyme NADH (the reduced form of nicotinamide adenine dinucleotide). In the laboratory, a similar function is performed by metal hydrides such as sodium borohydride and lithium aluminium hydride as well as aluminium isopropoxide (Meerwein-Ponndorf-Verley reduction).

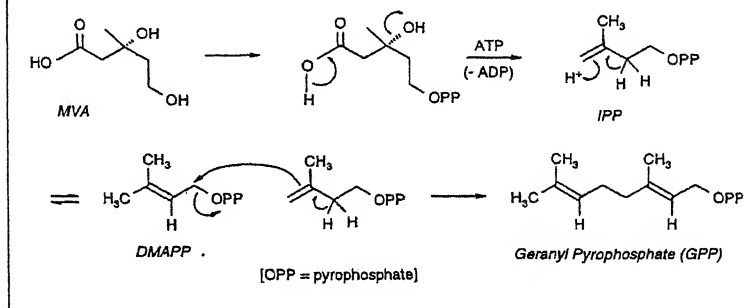
Box 5

In contrast to most laboratory reactions, biochemical reactions result in the formation of one of the enantiomorphs of an optically active compound. This is a consequence of the fact that these reactions are catalysed by enzymes which themselves are chiral in character.

Box 6

The bicyclic monoterpenoids produced from this carbocation include, besides compounds having pinane and bornane structures, those derived from carane and thujane, each of which has a cyclopropane ring.

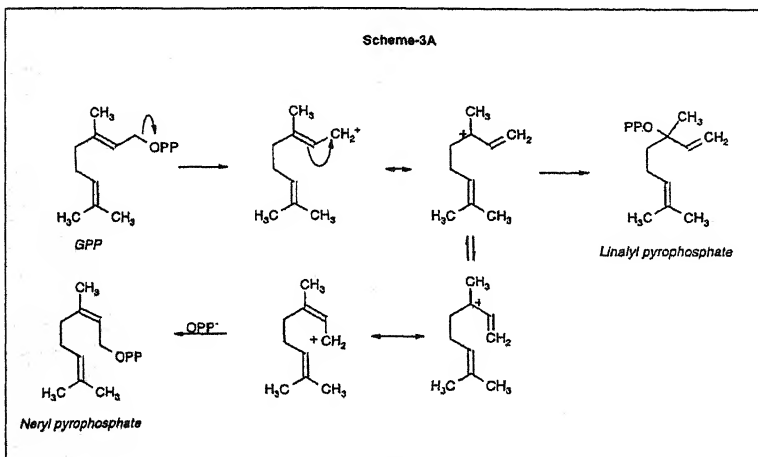
Scheme-2



Geranyl pyrophosphate can isomerise into neryl pyrophosphate and linalyl pyrophosphate (*Scheme-3A*). Such an isomerism occurs with ease since the pyrophosphate group is a good leaving group and is also a nucleophile. Linalyl pyrophosphate loses the phosphate grouping with the simultaneous interaction of the two double bonds as shown in *Scheme-3B*. The result is a carbocation with a six membered ring. It is important to note that such a cyclisation requires the appropriate conformation of the acyclic precursor as indicated in the Scheme. This carbocation can stabilise itself in several possible ways, one of which is an intra molecular interaction between the cationic center and the endocyclic double bond. This reaction itself can occur in either of two ways, one leading to the pinene structure and the other to borneol as shown in the Scheme. Oxidation of borneol gives camphor. This series of reactions can be used to illustrate the nuances in nucleophilic substitution reactions and the generation and fate of carbocations. Figure out for yourself the other possible modes of stabilisation of the afore mentioned carbocation (*Box 6*).

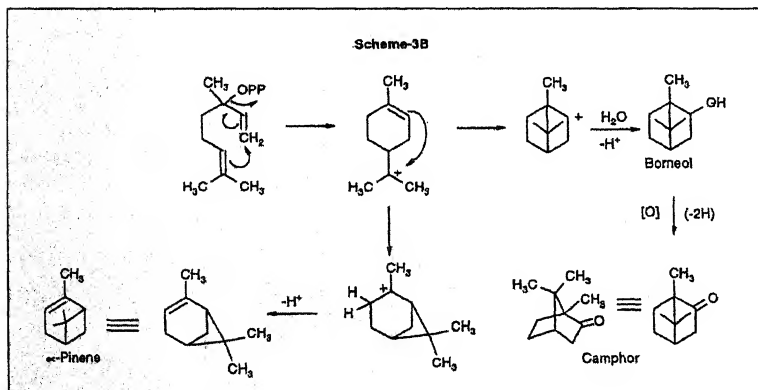
Komppa's synthesis of camphor was a classical example of a total synthesis designed for confirming a structure deduced from analytical and degradation studies. A key intermediate was camphoric acid which was synthesised starting from diethyl oxalate. The sequence involved several well known and

Scheme-3A



unambiguous reactions such as Claisen condensation, base catalysed C-methylation and Meerwein-Ponndorf-Verley reduction. We will not discuss this synthesis here, since a description of this can be found in several text books of organic chemistry. We will, instead, describe three newer and shorter methods of synthesis. In one of them the primary target molecule is camphene which is then converted to camphor via isoborneol (Scheme-4). The first step is a Diels-Alder reaction (Box 7) between cyclopentadiene and mesityl oxide. Subsequent steps involve catalytic hydrogenation, a haloform reaction, reduction of carbonyl to CH_2OH group and dehydration via the tosyl ester. Camphene (Box 8) thus obtained is converted into camphor in two steps.

Scheme-3B



Box 7

The Diels-Alder reaction is one of the most popular among preparative organic reactions as it is stereospecific. It is a thermally allowed cyclo-addition reaction between a diene and a dienophile and comes under the broad class of pericyclic reactions. The Nobel laureates Woodward, Hoffmann and Fukui, put forward a theoretical framework for understanding the mechanism of this and other pericyclic reactions.

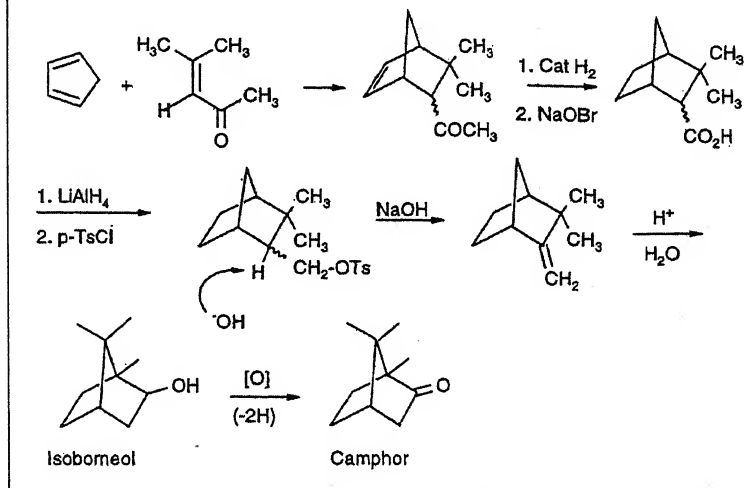
Box 8

(-) Camphene is a crystalline bicyclic monoterpene hydrocarbon which was first isolated in 1888 from the essential oil of *Abies sibirica*. It can also be obtained from α -pinene and can be converted in turn, into camphor via isobornyl acetate. In these transformations, Wagner-Meerwein rearrangements play a major role.

Box 9

A regiospecific reaction is one in which one single structural isomer is formed exclusively. The reaction under discussion is not regiospecific as both the possible structural isomers are obtained. From the preparative point of view, a regiospecific reaction is to be preferred to a non-regiospecific reaction.

Scheme-4

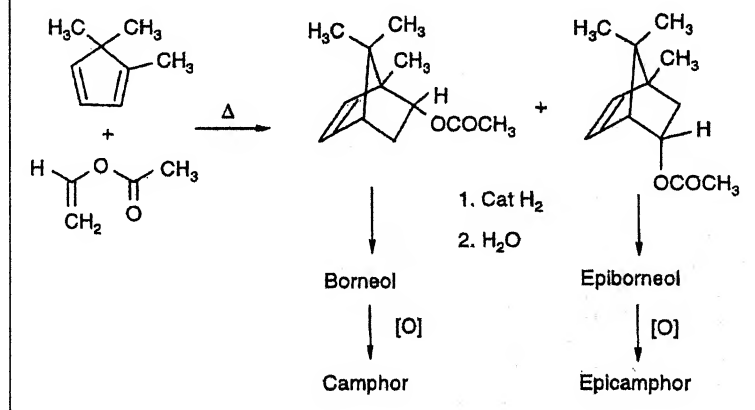


A Diels-Alder reaction is the first step in another direct synthesis of borneol wherein the starting materials are 1,5,5-trimethylcyclopentadiene and vinyl acetate. However, this reaction is not regiospecific (*Box 9*) and a mixture of bornyl and epibornyl acetates is obtained. Hydrolysis and oxidation yield camphor and epicamphor as shown in *Scheme-5*.

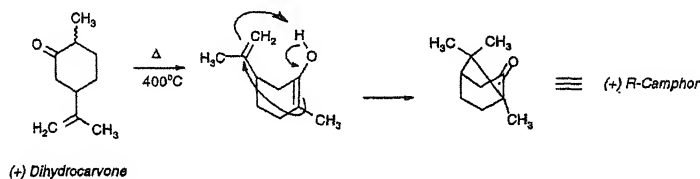
Box 10

An asymmetric synthesis is one wherein one enantiomorph of a chiral compound is formed either exclusively or as the major product. The reaction described here is a good example of an asymmetric synthesis as the product is 80% optically pure.

Scheme-5



Scheme-6



The third synthesis is an asymmetric synthesis (*Box 10*) in which (+) camphor of 80% optical purity is produced by heating (+)-dihydrocarvone at 400°C . The enol form of the ketone undergoes an ene reaction as shown in *Scheme-6*. Try to understand the mechanism of each step of the above synthetic sequences as it is the key to the understanding of organic reactions. Camphor is not a particularly complex molecule but the syntheses described herein do have ingenuity and an elegant architectural planning.

Address for Correspondence

N R Krishnaswamy
No.12, 9th Main Road
Banashankari 2nd Stage
Bangalore 560 070, India

It is necessary to emphasise that in these six articles we have only touched the fringes of the chemistry of natural products. But in doing so, we hope we have been able to give you a fairly clear overall picture. Conventionally, natural product chemistry is taught as a (large) chapter within organic chemistry. The reality, however, is that it is organic chemistry (with an addendum trailing into biology) magnified several times. All aspects of organic chemistry with all their nuances can be taught and learnt using natural products as illustrative examples. This has been the main objective of this series of articles. Learning (any subject) is largely a personal experience and the major function of a good book is to provide the right kind of framework, incorporating factual material with a sense of priority as to what is more important and what is less. A book has its own limitations, the major restricting factor being space. This constraint is more pronounced in an article appearing in a journal. A good teacher fortifies what she/he gets from a reliable book with additional material gathered from other sources, edits them to weed out trivialities, and translates this impersonal, very often 'dry' (but accurate) parcel of information into something lively, attractive and palatable. Her (his) main function is to instill a craving for knowledge and stimulate thinking and reasoning power in her (his) students. If a good book is comparable to a Shakespearean play (for example, *Hamlet*), a teacher's performance in the class room must resemble interpretation by someone like Sir Laurence Olivier! Those of us who have had the good fortune to learn the chemistry of natural products under outstanding masters know that the subject is no less enjoyable than a Shakespearean drama. We do not know how far we have succeeded in conveying to our readers the beauty and richness of the subject through these six articles. We would like to hear from students who have read these articles.

Fourier Series

The Mathematics of Periodic Phenomena

S Thangavelu



The author received his Ph.D from Princeton University in 1987. After spending several years at TIFR Centre, Bangalore, he is currently with the Indian Statistical Institute, Bangalore. His research interests are in harmonic analysis and partial differential equations.

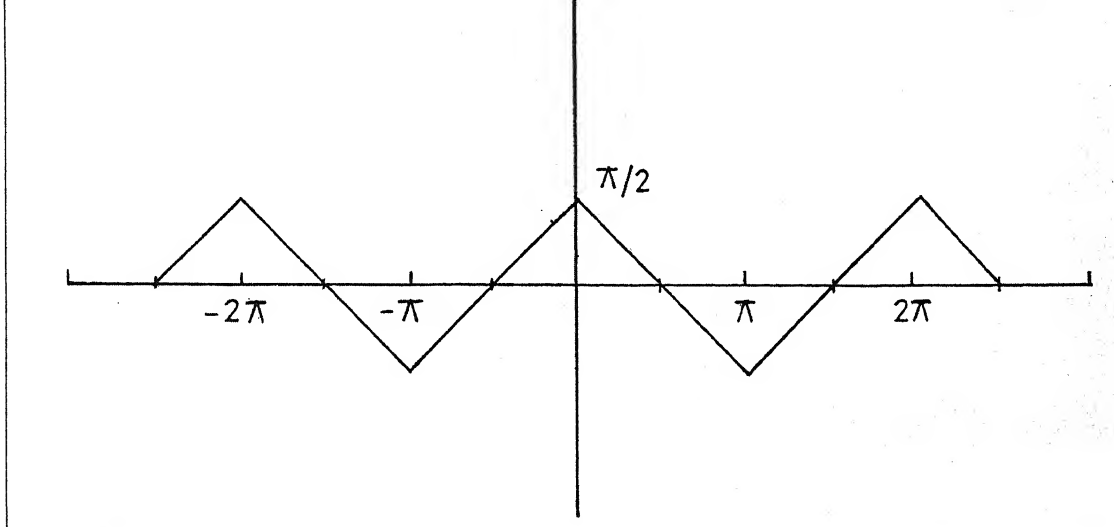
There are several natural phenomena that are described by periodic functions. The position of a planet in its orbit around the sun is a periodic function of time; in chemistry, the arrangement of molecules in crystals exhibits a periodic structure. The theory of Fourier series deals with periodic functions. By a periodic function we mean a function f of a real variable t which satisfies the relation $f(t + c) = f(t)$ for all t , where c is constant. Here c is called a *period* of f . The simplest examples of periodic functions of period 2π are provided by the trigonometric functions $\sin kt$ and $\cos kt$ for integer k . It follows that linear combinations of the form

$$P_n(t) = \sum_{k=0}^n (a_k \cos kt + b_k \sin kt)$$

are also periodic with the same period 2π . These are called *trigonometric polynomials*. The basic idea behind Fourier series is that any periodic function (of period 2π) can be approximated by trigonometric polynomials. (Notice that it is enough to consider periodic functions with period 2π because a function f with some other period, say c , can be converted to a function of period 2π by a suitable change of variable. (Exercise !)) The subject of Fourier series finds a wide range of applications from crystallography to spectroscopy. It is one of the most powerful theories in the history of mathematics and has stimulated the development of several branches of analysis.

Consider the function $f(t)$ defined by the formula

$$f(t) = \begin{cases} \frac{\pi}{2} - t, & 0 \leq t \leq \pi \\ \frac{\pi}{2} + t, & -\pi \leq t \leq 0 \end{cases}$$



which for other values of t is extended by periodicity. Thus the graph of the function looks as shown in *Figure 1*.¹

This function is continuous but not differentiable at the points $0, \pm\pi, \pm2\pi, \dots$ where the graph has corners. The function $f(t)$ is even. The functions $\cos kt$ are all even and the functions $\sin kt$ are all odd.² Hence we may expect to approximate f by trigonometric polynomials of the form $\sum_{k=0}^l a_k \cos kt$. In fact consider

$$P_n(t) = \frac{4}{\pi} \sum_{k=0}^n (2k+1)^{-2} \cos(2k+1)t.$$

The graphs of these functions for $n=2$ and for $n=4$ are shown in *Figure 2*.

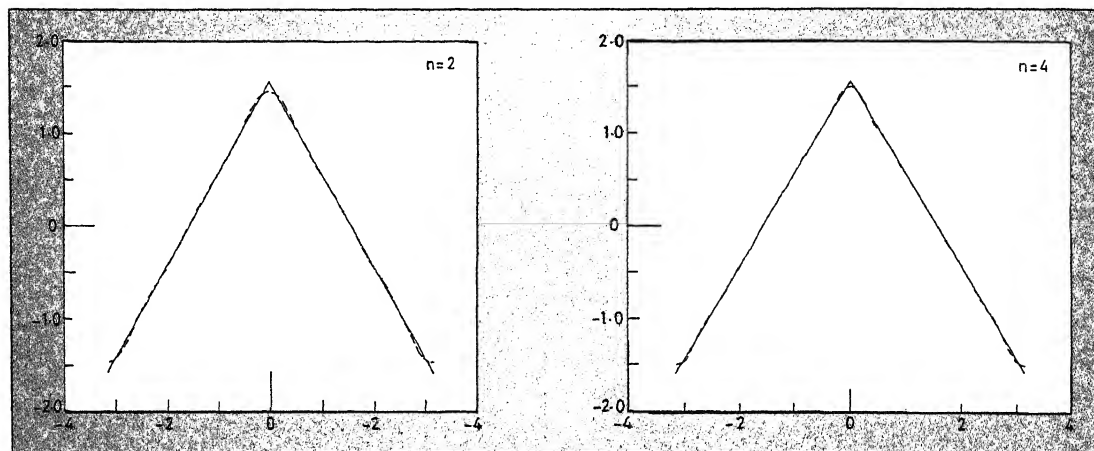
From the picture it is clear that even for small values of n , the trigonometric polynomial P_n approximates f very well.

Now you may be wondering why we have chosen the particular trigonometric polynomials P_n in order to approximate f . As we have already remarked the absence of the sine terms may be

¹ The author thanks A S Vasudevamurthy for helping him with the various graphs in this article. The material in some of the boxes and footnotes were provided by C Varughese.

² A function h is *even* if $h(x) = h(-x)$ for all x and *odd* if $h(x) = -h(-x)$ for all x .

Figure 2



explained by the fact that f is even and $\sin kt$ are all odd; but why did we choose the coefficients $(2k+n)^{-2}$ for the terms $\cos(2k+1)t$? Is there a way of relating the function f to the coefficients $(2k+n)^{-2}$? To answer all these questions, let us go back in history by a couple of centuries. The basic idea that any periodic function may be expanded in terms of trigonometric polynomials is attributed to D Bernoulli. This Swiss mathematician believed in the 'universal character' of the trigonometric polynomials much before Fourier. It all started with his works in 1747, 1748 and 1753 where he studied the vibrating string problem: find the solutions of the following partial differential equation with the given initial conditions

$$\frac{\partial^2 u(x, t)}{\partial t^2} = \frac{\partial^2 u(x, t)}{\partial x^2}$$

$$u(x, 0) = f(x), \quad u(-\pi, t) = u(\pi, t) = 0$$

This is called the *wave equation* and the graph of $u(x, t)$ represents the shape of the vibrating string at time t . The initial shape of the string is given by the function f and the condition $u(-\pi, t) = 0 = u(\pi, t)$ means that the string is kept fixed at the ends.



This equation had been studied by Euler and D'Alembert before Bernoulli. D'Alembert gave the solution of the above problem in the form

$$u(x, t) = \frac{1}{2}(f(x+t) + f(x-t)).$$

We can easily check that this function u is indeed a solution provided f is an odd periodic function with period 2π and is twice differentiable. But Bernoulli had a different idea. The functions

$$u_k(x, t) = a_k \cos kt \sin kx$$

satisfy the wave equation and following physical ideas Bernoulli suggested solutions of the form

$$u(x, t) = \sum_{k=0}^{\infty} a_k \cos kt \sin kx.$$

Based on this observation Bernoulli was led to believe in the possibility of expanding arbitrary periodic f with $f(\pi) = f(-\pi) = 0$ in terms of $\sin kx$ but he did not have a clue as to how to calculate the various coefficients!

Mathematics had to wait for almost fifty years for a formula to calculate the *Fourier coefficients*. In 1807, the French mathematician Jean Joseph Fourier working on the heat equation found a way to calculate the coefficients.

The idea of Fourier is every simple. Let us rewrite the trigonometric polynomials

$$P_n(t) = \sum_{k=0}^n (a_k \cos kt + b_k \sin kt)$$

in the form

$$P_n(t) = \sum_{k=-n}^n \alpha_k e^{ikt}.$$

That these two are equivalent follows from the well known relation



$$e^{ikt} = \cos kt + i \sin kt, \text{ where } i = \sqrt{-1}.$$

Suppose that these trigonometric polynomials $P_n(t)$ converge to $f(t)$ in any reasonable sense so that we can take the limit under the integral sign to get

$$\lim_{n \rightarrow \infty} \int_{-\pi}^{\pi} P_n(t) e^{-imt} dt = \int_{-\pi}^{\pi} f(t) e^{-imt} dt, \text{ for each fixed } m.$$

The integrals on the left hand side of the above equation can be easily calculated. In fact

$$\int_{-\pi}^{\pi} P_n(t) e^{-imt} dt = \sum_{k=-n}^n \alpha_k \int_{-\pi}^{\pi} e^{i(k-m)t} dt$$

and an easy integration reveals that

$$\int_{-\pi}^{\pi} e^{i(k-m)t} dt = 0$$

whenever $k \neq m$. And when $k=m$ we get 2π for the value of the integral. Thus we have the formula for α_m :

$$\alpha_m = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e^{-imt} dt.$$

To emphasize that α_m depends on f we write $\alpha_m = \hat{f}(m)$. The associated series

$$f(t) = \sum_{k=-\infty}^{\infty} \hat{f}(k) e^{ikt}$$

is called the *Fourier series* of the function f .³ $\{\hat{f}(k)\}_{k=-\infty}^{\infty}$ are called the *Fourier coefficients*.

(Exercise: If f is differentiable and f' is continuous show that the Fourier series corresponding to f' is given by

$$\sum_{k=-\infty}^{\infty} (ik) \hat{f}(k) \exp(ikt).)$$

Fourier's idea is so simple that one may wonder why it eluded D'Alembert, Bernoulli, and even the great Euler. But one has to

³ The equality here must be taken to mean that the partial sums of the series on the right approximate the function f in a sense that will be described later. The two sides are not necessarily equal for every value of t , although it is customary to write it in this form.

bear in mind that these mathematicians of the seventeenth century had to deal with concepts like functions, integrals and the convergence of infinite series when they were not well understood. For Euler, a function always meant an analytic expression; Bernoulli believed that any possible position of a vibrating string represents a function and can be given by an analytic expression. It was Fourier who went one step further to consider all functions either continuous or discontinuous and boldly asserted that they can be represented by an infinite series which came to be known as Fourier series. The moral of the story is: a simple argument can still be very deep and it may not be very simple to discover!

Let us go back and calculate the Fourier coefficients of the function f which we started with. As f is even

$$\begin{aligned}\hat{f}(0) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) dt \\ &= \frac{1}{\pi} \int_0^{\pi} \left(\frac{\pi}{2} - t \right) dt = 0\end{aligned}$$

For $k \neq 0$,

$$\begin{aligned}\hat{f}(k) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e^{-ikt} dt \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \cos kt dt\end{aligned}$$

since the integral

$$\int_{-\pi}^{\pi} f(t) \sin kt dt = 0$$

on account of the fact that $\sin kt$ is odd. Therefore

$$\begin{aligned}\hat{f}(k) &= \frac{1}{\pi} \int_0^{\pi} \left(\frac{\pi}{2} - t \right) \cos kt dt \\ &= -\frac{1}{\pi} \int_0^{\pi} t \cos kt dt\end{aligned}$$

since the integral $\int_0^\pi \cos kt \, dt = 0$. Integrating by parts in the last integral we can easily see that

$$\hat{f}(k) = \frac{1}{\pi k^2} (1 - \cos k\pi) = \frac{1}{\pi k^2} (1 - (-1)^k),$$

the last equality holds because $\cos k\pi = (-1)^k$. Thus $\hat{f}(2k) = 0$ and $\hat{f}(2k+1) = \frac{2}{\pi} (2k+1)^{-2}$ and consequently the Fourier series of f becomes

$$f(t) = \frac{2}{\pi} \sum_{k=-\infty}^{\infty} (2k+1)^{-2} e^{i(2k+1)t}$$

Combining the terms corresponding to $(2k+1)$ and $-(2k+1)$ we get

$$f(t) = \frac{4}{\pi} \sum_{k=0}^{\infty} (2k+1)^{-2} \cos(2k+1)t.$$

Now, it should be clear why the polynomials $P_n(t)$ gave good approximations to the function $f(t)$!

Thanks to Fourier, we now have a way of writing down the Fourier series of any periodic function. But what is the guarantee that the Fourier series actually represents the function? In other words how do we know if the partial sums defined by

$$S_n f(t) = \sum_{k=-n}^n \hat{f}(k) e^{ikt}$$

converge to the function f pointwise, i.e. how do we know that $S_n f(t)$ converges to $f(t)$ as $n \rightarrow \infty$, for every t ? It is not very hard to prove that for the particular function of *Figure 1*, we have such convergence. *Figure 2* is convincing evidence of this fact. Fourier believed and asserted that, in general, $S_n f$ always converges to the function f , but he offered no proof. *Fourier's assertion turned out to be wrong as we will see shortly*. This by no means diminishes the greatness of Fourier's contributions. In fact, with a different interpretation of convergence his assertion is correct: the Fourier



Box 1

If we do not insist on pointwise convergence, but are willing to settle for *mean square convergence*, then indeed for a wide class of 2π -periodic functions f , the Fourier series converges to f . All we need to assume is that f is *square-integrable* i.e. $\int_{-\pi}^{\pi} |f(x)|^2 dx$ is finite. For such f , $S_n f$ converges to f in the mean-square sense even if for *many* values of t , $S_n f(t)$ does *not* converge to $f(t)$ i.e. $\frac{1}{2\pi} \int_{-\pi}^{\pi} |S_n f(t) - f(t)|^2 dt \rightarrow 0$ as $n \rightarrow \infty$ i.e. the *average value* of the square of the error $|S_n f(t) - f(t)| \rightarrow 0$ as $n \rightarrow \infty$. If f represents a periodic signal, physicists and electrical engineers will tell you that $\int_{-\pi}^{\pi} |f(t)|^2 dt$ represents the *energy* in the signal. Using the fact about mean square convergence explained above, we can actually prove that $\int_{-\pi}^{\pi} |f(t)|^2 dt = 2\pi \sum_{k=-\infty}^{\infty} |\hat{f}(k)|^2$. However instead of using $\exp(ikt)$, if we use $\sin kt$ and $\cos kt$ and write the Fourier series as $a_0 + \sum_{k=1}^{\infty} b_k \cos kt + \sum_{k=1}^{\infty} c_k \sin kt$, then the above formula can be written as $\int_{-\pi}^{\pi} |f(t)|^2 dt = 2\pi |a_0|^2 + \pi \sum_{k=1}^{\infty} |b_k|^2 + \pi \sum_{k=1}^{\infty} |c_k|^2$ (why?). Finally if g is another 2π -periodic function with corresponding Fourier series $A_0 + \sum_{k=1}^{\infty} B_k \cos kt + \sum_{k=1}^{\infty} C_k \sin kt$, then one has the more general formula $\int_{-\pi}^{\pi} f(t)g(t)dt = 2\pi a_0 A_0 + \pi \sum_{k=1}^{\infty} b_k B_k + \pi \sum_{k=1}^{\infty} c_k C_k$.

series of a function converges in what the modern day mathematicians call *the sense of distributions*.

The theory of distributions, an important development of this century, was actually anticipated by Fourier as one can see from his treatment of Fourier integrals! For another mode of convergence, which is valid for a very wide class of functions, see *Box 1*.

The pointwise convergence of the partial sums $S_n f$ to the function f fails — it fails dramatically at some points as we see from the following example. Consider the function $g(t) = -1$ for $-\pi \leq t \leq 0$, $g(t) = 1$ for $0 < t \leq \pi$. This function is discontinuous at



the points $0, \pm\pi, \pm2\pi, \dots$, where it has a jump of size 2. We can easily calculate the Fourier series of this function. We leave it as an exercise to the reader to show that $\hat{g}(0) = 0, \hat{g}(2k) = 0$ and that $\hat{g}(2k+1) = -i \frac{2}{\pi} (2k+1)^{-1}$. Therefore, the Fourier series of g takes the form

$$g(t) = \frac{-2i}{\pi} \sum_{k=-\infty}^{\infty} (2k+1)^{-1} e^{(2k+1)it}$$

or combining the terms with $(2k+1)$ and $-(2k+1)$ we can write it as

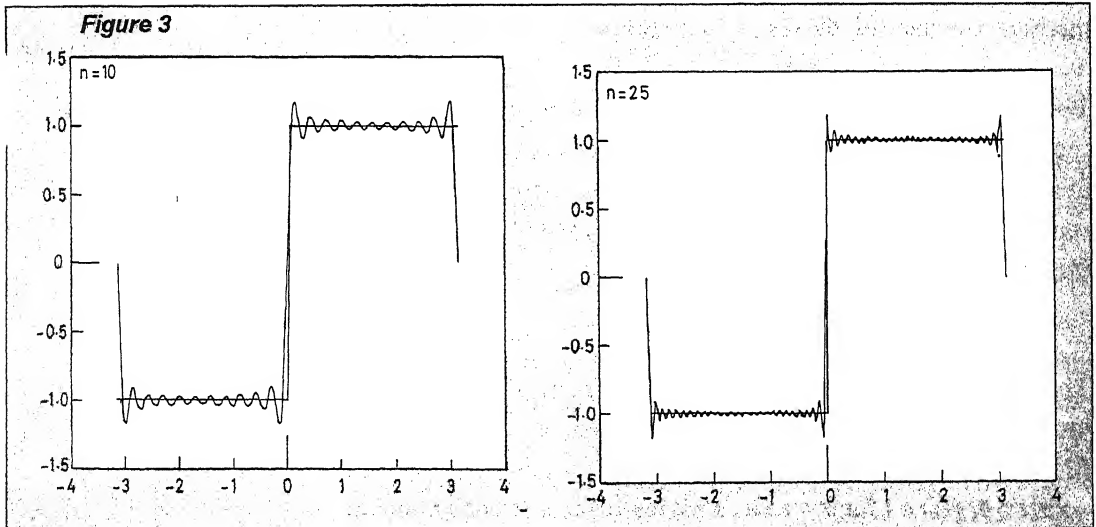
$$g(t) = \frac{4}{\pi} \sum_{k=0}^{\infty} (2k+1)^{-1} \sin(2k+1)t.$$

Consider now the partial sums associated to the above Fourier series:

$$S_n g(t) = \frac{4}{\pi} \sum_{k=0}^n (2k+1)^{-1} \sin(2k+1)t.$$

The graphs of some of these trigonometric polynomials are shown in *Figure 3*.

From the graphs we observe that $S_n g$ approximates g well except in a small neighbourhood of $t=0$ where it overshoots and under-



shoots the levels ± 1 respectively. Naturally one expects that the overshoot and undershoot tends to zero as n goes to infinity; but surprisingly this does not take place. Instead, they are always about 9%. This is called the *Gibbs phenomenon* because it was pointed out by Gibbs in a letter to *Nature* in 1899. The letter was a reply to one from Michelson (of the *Michelson - Morley experiment* fame), who apparently was angry with the machine he used for computing Fourier series, as it failed badly at jumps!

The Gibbs phenomenon that occurs in the vicinity of jump discontinuities can be proved rigorously. This already shows that we cannot expect the Fourier series to converge at all points. One may think that the situation will be different in the case of continuous functions. But already in 1876, du Bois - Reymond constructed a continuous function whose Fourier series fails to converge at a given point. Several eminent mathematicians including Dirichlet, Riemann and Cantor occupied themselves with the problem of convergence of Fourier series. Many positive results were proved in the course of time but in 1926 the Russian

Box 2

Let $f(t) = t$, $-\pi \leq t < \pi$, and extended for other values of t by periodicity. Use the result at the end of *Box 1* to calculate the value of π^2 as the sum of an infinite series.

Box 3

Notice that the functions $\sin kt$ and $\cos kt$ are periodic with period $(2\pi/k)$. Of course they are periodic with period 2π but they are also periodic with the smaller period $(2\pi/k)$ and this is the smallest period for these functions. Thus, they have *frequency* $(k/2\pi)$. By choosing various values of k , we get what we can think of as a *standard family* of signals. If f is a periodic signal (with period 2π), its Fourier series is a decomposition of f into signals from the standard family. The relation $\int_{-\pi}^{\pi} |f(t)|^2 dt = 2\pi \sum |\hat{f}(k)|^2$ (see *Box 1*) is then a description of how the energy contained in the signal f is distributed among various frequencies. The relative amount of energy contained at any frequency is determined by the square of the modulus of the corresponding Fourier coefficient.



mathematician A N Kolmogorov came up with a discouraging result. He constructed an integrable (in the sense of Lebesgue) function whose Fourier series diverges everywhere ! After this example by Kolmogorov, mathematicians started believing that one day someone will construct a continuous function with everywhere divergent Fourier series. That day never came. Instead, in 1966, the Swedish mathematician L Carleson showed that in the case of continuous functions the Fourier series can fail to converge only on a set which is *negligible* in a certain sense (i.e. of *measure zero*)!

When the function f is smooth, say once differentiable and f' is continuous, it can be shown that the Fourier series converges to the function at each point, and even *uniformly*. But in applications we have to deal with functions having very bad discontinuities. Dirichlet and Dini found local conditions on the functions that will ensure pointwise convergence of the Fourier series. These results are adequate for the purpose of applications. However, mathematicians not being content with such results have been seeking more and more general results proving the convergence of Fourier series for larger and larger classes of functions! They have also worked with different notions of convergence although unfortunately many of these cannot be explained here without assuming a lot of Lebesgue theory of integration.

We would like to conclude this article with the following result of Fejer which treats the class of continuous functions as a whole. As we know, given any point t_0 there is a function in this class whose Fourier series diverges at that point. In 1904, the Hungarian mathematician Fejer had the brilliant idea of considering the *averages of partial sums instead of the partial sums themselves*. Thus he considered

$$\sigma_n f = \frac{1}{n+1} (S_0 f + S_1 f + \cdots S_n f).$$

These are called the *Cesaro means* and it is easy to see that they can be written in the form

$$\sigma_n f(t) = \sum_{k=-n}^n \left(1 - \frac{|k|}{n+1}\right) \hat{f}(k) e^{ikt}.$$

The celebrated theorem of Fejer says that when f is continuous the Cesaro means $\sigma_n f$ converge to f , not only pointwise, but also in the stronger sense of *uniform convergence*. Hence trigonometric polynomials are *dense* in the class of continuous functions !

The body of literature dealing with Fourier series has reached epic proportions over the last two centuries. We have only given the readers an outline of the topic in this article. For the full length episode we refer the reader to the monumental treatise of A Zygmund. Beginners will find the books of Bhatia and Körner very helpful. For more advanced readers, the books of Dym - McKean, Folland, Helson and Katznelson make enjoyable reading. The article of Gonzalez - Velasco deals with the influence of Fourier series on the development of analysis and we highly recommend it.

Suggested Reading

- ◆ A Zygmund. *Trigonometric series*. Volumes I and II. Cambridge University Press, 1959.
- ◆ H Dym and H P McKean. *Fourier series and integrals*. Academic Press, 1972.
- ◆ Y Katznelson. *An introduction to harmonic analysis*. Dover Publications, 1976.
- ◆ T W Körner. *Fourier Analysis*. Cambridge University Press, 1989.
- ◆ G B Folland. *Fourier analysis and its applications*. Wadsworth and Brooks, 1992.
- ◆ E Gonzalez-Velasco. *Connections in mathematical analysis, the case of Fourier series*. American Mathematical Monthly. May, pp 427-441, 1992.
- ◆ R Bhatia. *Fourier series*. TRIM-2. Hindusthan Book Agency, 1993.
- ◆ H Helson. *Harmonic analysis* (2nd edition). TRIM-7. Hindusthan Book Agency, 1995.

Address for Correspondence
S Thangavelu
Department of Mathematics
and Statistics, University of
New Mexico, Humanities
Building 419, Albuquerque,
NM 87131-1141, USA

Barbara McClintock and the Discovery of Jumping Genes

Vidyanand Nanjundiah



Vidyanand Nanjundiah works in the Developmental Biology and Genetics Laboratory at the Indian Institute of Science. After a Master's degree in physics he took up biology. He is interested in evolutionary biology and pattern formation during development.

Barbara McClintock's life shows us how important it is to nurture original and unconventional thinking in science if we are to get out of the rut of ordinariness. After a long period of relative neglect, she was awarded the Nobel Prize in 1983 for her work on genetic instability (transposition).

The history of modern genetics begins with the experiments of Gregor Mendel (1822-1884). Mendel found that when hereditary traits were followed through successive generations of hybridisation, the numbers of offspring that resembled parental types were in simple numerical ratios relative to one another – 1:1, or 3:1, or 9:3:3:1, and so on. The most straightforward explanation of these numbers was that the traits were associated with discrete, indivisible entities, later to be called genes. Mendel's observations languished in obscurity for 34 years until their rediscovery in 1900. Following this, rapidly accumulating data enabled genes to be mapped. Genes were found to be organised into distinct groups that were arranged in a linear order. This, along with other information, suggested that genes kept company with the thread-like structures called chromosomes that existed inside the cell's nucleus. Over a period of about 50 years biologists dealt with genes and built up a successful predictive science of genetics. All the same, no one knew for sure what genes were made of (a situation that has interesting parallels to the development of the atomic theory). An understanding of the nature of the gene had to await the identification of DNA as the genetic material.

The final piece of evidence for a physical model of genes emerged from two independent experiments whose results were published

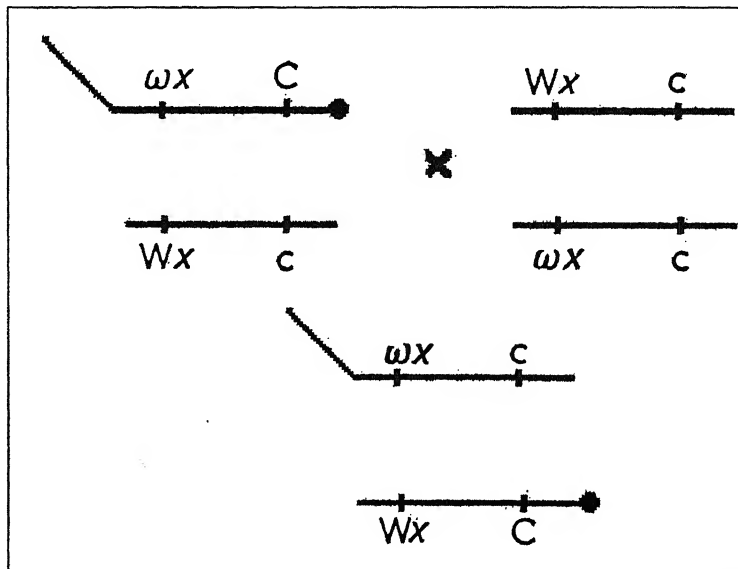


in 1931. Harriet Creighton and Barbara McClintock, working in the U.S.A. with maize, and Curt Stern, working in Germany with the fruit fly *Drosophila*, finally proved that genes were associated with chromosomes. Their conclusion was based on the observation that when genes appeared to 'cross over' from one genetic neighbourhood to another, so did the chromosomal material. Immediately acclaimed as a landmark in classical genetics, the finding made McClintock famous. Having thereby endowed genes with a measure of solidity, so to speak, she next went on to do just the opposite. She developed the unsettling idea that genes could be unstable, an insight that ought to have caused a sensation. Instead, it aroused indifference. This state of affairs lasted for some 30 years. Then a rush of independent discoveries brought genetic instability, and with it McClintock, back into the limelight.

Born in 1902 as the third of four children, Barbara McClintock was a fun-loving but solitary child as she grew up in New York. She had unconventional parents to whom what their daughter *could* be was more important than what she *should* be. Parental support and understanding continued in the face of evident oddities: it was apparent that young Barbara was not cut out to be part of the mainstream. Her pursuits were scholarly but at the same time individualistic. She joined Cornell University in 1918 after finishing high school and, in a step that would prove momentous for the future, became interested in the chromosomal organization and genetics of maize while still an undergraduate. Her first major contribution to the field was to identify each of the 10 chromosomes that maize has. Later she was to order several maize genes using a method now known as deletion mapping. Along with parallel work by others on *Drosophila*, this was an exciting development. McClintock became the intellectual driving force of an extraordinarily talented maize genetics group that was assembled at Cornell by R Emerson; among others the group included George Beadle who was also to win a Nobel prize later. A PhD degree in 1927 was followed by the famous paper with Creighton which provided a materialistic grounding to



Figure 1 Creighton and McClintock made use of two genetic markers on chromosome number 9 of maize. One gene affected seed coat colour (C coloured, c colourless) and the other affected the composition of the food reserve (Wx = starchy, wx = waxy). C and Wx are dominant, meaning that they exert their effects in single doses; c and wx are recessive and have to be present in two doses in order to be perceived. A plant having the chromosomal combination ($wx C/Wx c$), with a knob on the chromosome containing C , was crossed with one having no knobs, two copies of c and one each of Wx and wx . Different types of progeny were produced; some possessed the coloured character in combination with starchy. As indicated in the lower portion of the figure, this was invariably associated with the exchange of that part of the chromosome containing the knob. (The bent portion is a fragment from another chromosome, also a visible marker like the knob.)



what had been, until then, the formal concept of a gene.

The demonstration of crossing over could not have been simple (Figure 1). Chromosome number 9 in maize had a variant form with a knob-like visible bulge at one end. Gene locations had already been mapped on the chromosome using conventional techniques. Creighton and McClintock carried out matings of the knobbed stock with the standard variety. In the resulting progeny the presence of the gene that mapped closest to the knob invariably was associated with the knob itself, an observation that made it compelling to think of genes as things on chromosomes. McClintock's reputation continued to grow thereafter but her career did not prosper. It did not prosper in the manner it might have, had she been more easygoing, more conventional and less forthcoming in displaying her intelligence. Getting accepted in the male world of science was not easy either. The awe and respect in which she was held by her peers contrasted starkly with her inability to make it in academia.

Eventually she left Cornell University: her position there was always tenuous and she was never made a member of the faculty. She went on to join the University of Missouri as an assistant

professor (1936-1941). Scientifically, this phase of her life was marked by the observation that broken ends of maize chromosomes behaved in unusual ways and that repairing the breaks required genetic activity. This observation was to come into its own much later, initially in connection with McClintock's work on transposable elements and more recently with the study of chromosome ends (telomeres). The turning point in a hitherto uncertain life came in 1941 with the offer of a research appointment to the Cold Spring Harbor Laboratory (in New York) of the Carnegie Institution of Washington. McClintock remained at Cold Spring Harbor for the rest of her life. (She died in 1992.) Here she was to be free from the imposition of applying for grants, free to pursue her inclinations and free to follow the unexpected.

Genetic instability as such was not the issue. The surprise was that certain genes could move from place to place thousands of times more frequently than the rate at which mutations were known to occur.

The work on genetic instability — or, to put it more correctly, genetic transposition, — for which she was to be honoured in 1983 with a Nobel Prize, dates from her early days in Cold Spring Harbor (*Figure 2*). Considering that this was a major scientific discovery, the fact that she was already in her 40s when she made it is unusual. There are other noteworthy aspects to it too. For one thing, it was not merely that she found something new, but that what she found turned conventional thinking upside down. (It should be remembered that gene mutations were known and accepted, so genetic instability as such was not the issue. The surprise was that certain genes could move from place to place thousands of times more frequently than the rate at which mutations were known to occur.) Secondly, as indicated at the beginning, the response to her announcement was a mixture of bafflement and silence. The more perceptive geneticists realized that something startling had emerged. But because it flew in the face of supposedly proven facts, transposition was relegated to the status of a brilliant discovery that was simultaneously a bizarre curiosity.

Because her observations flew in the face of supposedly proven facts, McClintock was in the unusual position of being admired



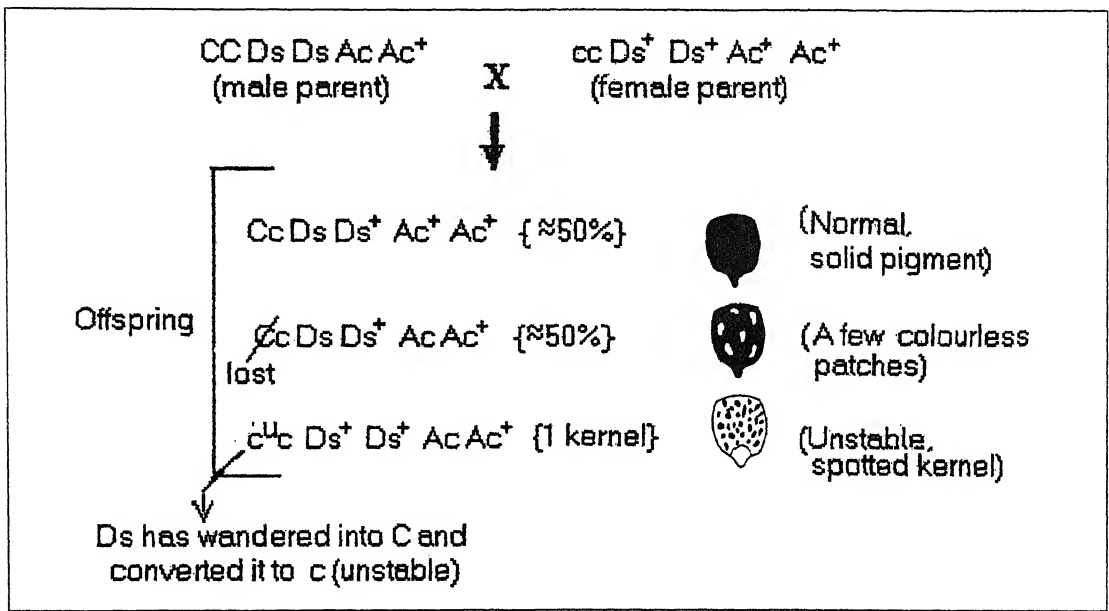


Figure 2 Activator (Ac) and Dissociator (Ds) are two unstable genetic elements discovered by McClintock (plus signs indicate the absence of the corresponding elements). The gene C gives rise to coloured kernels and c (in two copies) to colourless ones. About half the progeny of the genetic cross indicated at the top are normal-looking because in the absence of Ac, Ds is stable. In the remaining half, Ds caused a chromosomal break and loss of C, resulting in some colourless patches. But in one exceptional kernel, Ds wandered into C and converted it into an unstable colourless variant. This accounts for the numerous patches.

for her intellectual powers without being understood. Nor did it help that her papers were written in a style that made them next to impossible to grasp by anyone who had not already reached her own acuteness of perception with regard to the intricacies of maize genetics and cytology. To top it all, her revolutionary finding was not reported in the standard scientific literature, but instead in the annual reports of the Carnegie Institution or in Conference Symposium volumes. Behind this odd choice seems to have been the growing feeling on the part of McClintock that she had moved so far outside accepted modes of thinking that the establishment would not have been receptive to her anyway.

Nevertheless, she felt frustrated by her lack of success in communicating the excitement that she felt. According to her biographer, the reactions that McClintock's work elicited constituted a bitter anticlimax. (An indication of the general level of indifference is that even as late as 1965, there was no mention of genetic transposition in A H Sturtevant's otherwise authoritative history of genetics. Joshua Lederberg, the co-discoverer of genetic exchange in bacteria, is said to have come back from a meeting with her and said "By God, that woman is

either crazy or a genius".) On looking back at this continued neglect, one cannot help but ask disturbing questions related to the supposed impersonality, objectivity and receptivity of scientists. Other questions, also not easy to answer, have to do with science as it is practised in theory and in fact and the manner in which the scientific community reacts to deviations from accepted points of view.

The general appreciation of her work began to change with the arrival of the molecular biology revolution in the late 1960s. Soon it became apparent that the tendency of segments of DNA to move from one location to another was, if not universal, quite common. Interest in the field grew rapidly and it became apparent that mobile genetic elements, or 'jumping genes', as they came to be popularly known, were all over the place. One fact after another started emerging regarding their behaviour; it was demonstrated that comparable elements are involved in the transfer of resistance to antibiotics, in the generation of the immune response and in the spread of certain cancers by viruses. By the end of the 1970's the story of genetic instability, when set off against the long years of neglect, endowed McClintock's persona with a whiff of romance. Her status was raised almost to that of a cult figure.

Sure enough, even in what came to be thought of as 'her' field, genetic transposition, in one essential respect she remained an outsider. The question had to do with *why* genes jumped from place to place. All along, McClintock's stand had been that when segments of DNA moved, they did so for specific reasons. Because of their demonstrated ability to regulate the functioning of other pieces of DNA in maize, she coined the phrase 'controlling elements' to describe these segments. She went to the extent of advocating that similar controlling elements might be responsible for choreographing the orderly changes in patterns of gene expression that underly the development of a fertilized egg into an adult in plants and animals. Her viewpoint attracted virtually no support then, nor does it today. If anything there is an

Because her observations flew in the face of supposedly proven facts, McClintock was in the unusual position of being admired for her intellectual powers without being understood.

Mobile genetic elements are components of the dynamic structure of DNA.

increasing tendency to favour the hypothesis that transposable elements are parasitic molecules of DNA that have no purpose as such besides that of ensuring their own survival and reproduction. Other aspects of McClintock's thinking verged on the mystical, for instance her readiness to ascribe a certain kind of wisdom to cells and organisms. To be fair, this may have been no more than an overly metaphorical way of expressing a conviction that natural selection is all-pervasive, and that the demand of evolutionary adaptation has crafted living structures down to the smallest detail.

Today, the concept of mobile genetic elements is acknowledged as an essential component of what we have come to think of as the dynamic structure of DNA. To that extent McClintock's efforts have been vindicated. Her work is also a prime example of the way in which seemingly esoteric pure research can have unexpected offshoots. Her life provides a cautionary tale as well. The notion that the pursuit of knowledge for its own sake is a worthwhile enterprise in spite of its being fraught with uncertainty, is rapidly falling by the wayside. This is so even in supposed institutions of higher learning. As the years pass, it is unlikely that there will be many havens left of the sort that the Carnegie Institution provided to McClintock. If nothing else, contemporary standards of scientific assessment would have made it difficult for her to hold down a research job for any reasonable length of time. One needs only to visualise the dismay with which a committee would react to being told that someone had 30 publications in 'standard refereed journals' over half as many years, let alone over an entire career. Barbara McClintock's life shows us how important it is to nurture original and unconventional thinking in science if we are to get out of the rut of ordinariness, be it only occasionally.

Suggested Reading

- ◆ Evelyn Fox Keller. *A feeling for an organism, The life and work of Barbara McClintock*. W H Freeman & Company. New York, 1983.
- ◆ Fedoroff N V. Transposable genetic elements in maize. *Sci. Am.* June, 1984.

Address for Correspondence
Vidyanand Nanjundiah
Developmental Biology and
Genetics Laboratory
Indian Institute of Science
Bangalore 560012, India

What's New in Computers

Database Mining

J R Haritsa

Database Mining refers to the process of efficiently extracting patterns from huge historical databases. The pattern information helps organizations make better business decisions.

Introduction

Organizations typically collect vast quantities of data relating to their operations. For example, the Income Tax Department has several *terabytes* of data about taxpayers! In order to provide a convenient and efficient environment for users to productively use these huge data collections, software packages called *DataBase Management Systems* have been developed and are widely used all over the world (see *Box 1* for more details on database applications).

A DataBase Management System, or DBMS as it is commonly referred to, provides a variety of useful features: Firstly, business rules such as, for example, *the minimum balance to be maintained in a savings account is 100 rupees*, are not allowed to be violated. Secondly, modifications made to the database are never lost even if system failures occur subsequently. Thirdly, the data is always kept consistent even if multiple users access and modify the database concurrently. Finally, friendly and powerful interfaces are provided to ask questions on the information stored in the database.

Users of DBMSs interact with them in basic units of work called *transactions*. Transfer of money from one account to another, reservation of train tickets, filing of tax returns, entering marks



Jayant Ramaswamy Haritsa is with the Supercomputer Education and Research Centre and the Department of Computer Science and Automation at the Indian Institute of Science. He works on adapting database technology to new application domains such as VLSI design, flexible manufacturing, bio-diversity and real-time systems.

Box 1 Database Applications

Database Management Systems are used in a variety of applications all over the world. Typical applications and examples include the following:

Communications: Setting up and billing for telephone calls, electronic mail, data transfer, etc.

Finance: Banking, stock trading, merchandise purchases, etc.

Travel: Reservations and billing for airlines, hotels, cars, etc.

Manufacturing: Order entry, job and inventory planning and scheduling, accounting, etc.

Process Control: Control of machines, transporters, warehouses, etc.

Educational: Library search and reservation, reference material generation, etc.

Database Management Systems are meant to support applications that have to interface with large quantities of data. Typical sizes of databases are shown in the table below for a representative set of applications:

Application	Typical Size
Gene Mapping, Income Tax Returns, Voter Information	Terabytes (10^{12} bytes)
Banks, Libraries, Multimedia, Computer-Aided Design	Gigabytes (10^9 bytes)
Office Administration, Store Management, Timetables	Megabytes (10^6 bytes)

on a student's grade sheet, are all examples of transactions. A transaction is similar to the notion of a task in operating systems, the main difference being that transactions are significantly more complex in terms of the functionality provided by them (see *Box 2* for a more detailed description of transactions).

For large commercial organizations, where thousands of transactions are executed every second, highly sophisticated DBMSs such as DB2, Informix, Sybase, Ingres and Oracle are available and are widely used. For less demanding environments such as managing home accounts or office records, there are a variety of PC – based packages, popular examples being FoxPro, Microsoft Access and dBase IV.

Box 2 Transactions

The state of a database, at any given time, is defined to be the collection of items in the database and their current values. A database state is said to be consistent if the current values of the items in the database obey all the rules specified by the organization that is maintaining the database.

A transaction is the fundamental unit of database processing, which operates on the database and takes it from one state to another. It is a software module that is a sequence of actions with the following properties:

Atomicity: The changes made by the transaction to the database are atomic, that is, either all happen or none happen.

Consistency: The transaction is a correct program in that if it executes on a consistent database, the resultant database state is also guaranteed to be consistent.

Isolation: Although transactions are individually correct (by the Consistency property), unrestricted *concurrent* execution of transactions in a multiuser database system may lead to an inconsistent database state. Therefore, transactions regulate their execution in a manner such that, even though they execute concurrently, it appears as if each transaction executes in isolation from all other transactions.

Durability: After a transaction completes successfully, the changes it has made to the database state are durable. That is, these changes are not lost even if system failures occur subsequently.

Collectively, the four properties listed above are called *acid* properties! They are enforced by the DBMS software.

To illustrate the transaction concept, consider a university depositing a monthly scholarship into a student's account. This transaction is composed of two actions: debiting the amount from the university's account, and crediting the amount to the student's account. The transaction is atomic if it updates *both* the accounts. It is consistent if the credited amount is the same as the debited amount. It is isolated if the transaction can be unaware of other programs operating on the student's account concurrently (for example, the student making a withdrawal to pay his tuition fees). And it is durable if, once the transaction is complete, the balance in both accounts is guaranteed to reflect the transfer forever in the future.



In technical jargon, data patterns are called *rules* and the process of identifying such rules from huge historical databases is called *database mining*.

DataBase Management Systems typically operate on data that is resident on the hard disk. However, since disk capacities are usually limited, as more new data keeps coming in, organizations are forced to transfer their old data to tapes. Even if these tapes are stored carefully, the information in them is hardly ever utilized (similar to the files in our government offices!). This is unfortunate since historical databases often contain useful information, as described below.

Database Mining

The primary worth of the historical information is that it can be used to detect *patterns*. For example, a supermarket that maintains a database of customer purchases may find that customers who purchase coffee powder very often also purchase sugar. Based on this information, the manager may decide to place coffee powder and sugar on adjacent shelves to enhance customer convenience. The manager may also ensure that whenever fresh stocks of coffee powder are ordered, commensurate quantities of sugar are also procured, thereby increasing the company's sales and profits. Information of this kind may also be used beneficially in catalog design, targeted mailing, customer segmentation, scheduling of sales, etc. In short, the historical database is a 'gold mine' that can be profitably used to make better business decisions.

In technical jargon, data patterns (of the type described above) are called *rules* and the process of identifying such rules from huge historical databases is called *database mining*. This issue has recently become a hot topic of research in the database community and is the subject of this article. Readers may be aware that discovering rules from data has been an area of active research in artificial intelligence. However, these techniques have been evaluated in the context of small (in memory) data sets and perform poorly on large data sets. Therefore, database mining can be viewed as the confluence of machine learning techniques and the performance emphasis of database technology. In

particular, it refers to the efficient construction and verification of models of patterns embedded in large databases.

Rules

In normal usage, the term *rule* is used to denote implications that are always true. For example, Boyle's law i.e. $Pressure \times Volume = constant$ can be inferred from scientific data. For commercial applications, however, this definition is too restrictive and the notion of rule is expanded to denote implications that are *often*, but not necessarily *always*, true. To quantify this uncertainty, a *confidence factor* is associated with each rule. This factor denotes the probability that a rule will be true in a specific instance. For example, the statement "Ninety percent of the customers who purchase coffee powder also purchase sugar" corresponds to a rule with confidence factor 0.9.

For rules of the above nature to be meaningful, they should occur reasonably often in the database. That is, if there were a million customer purchases and, say, only ten of these customers bought coffee powder, then the above rule would be of little value to the supermarket manager. Therefore, an additional criterion called the *support factor* is used to distinguish *significant* rules. This factor denotes the fraction of transactions in the database that support the given rule. For example, a customer purchase rule with support factor of 0.20 means that twenty percent of the overall purchases satisfied the rule.

At first glance, the confidence factor and the support factor may appear to be similar concepts. However, they are really quite different: Confidence is a measure of the rule's strength, while support corresponds to statistical significance.

Formal Definition

Based on the foregoing discussion, the notion of a rule can be formally defined as $X \Rightarrow Y \mid (c, s)$, where X and Y are disjoint subsets of I , the set of all items represented in the database, c is

The confidence factor denotes the probability that a rule will be true in a specific instance. The support factor denotes the fraction of transactions in the database that support the given rule.



Table 1 Vegetable Purchase Database

Customer	Potatoes	Onions	Tomatoes	Carrots	Beans
1	N	N	N	N	Y
2	Y	Y	N	Y	Y
3	Y	Y	Y	N	N
4	Y	Y	N	N	Y
5	Y	Y	Y	N	N
6	Y	Y	Y	Y	N
7	Y	Y	Y	N	N
8	Y	N	N	N	Y
9	Y	Y	Y	N	N
10	Y	Y	Y	N	Y

the confidence factor, and s is the support factor. The factors, which are ratios, are usually expressed in terms of their equivalent percentages.

To make the above definition clear, consider a vegetable vendor who sells potatoes, onions, tomatoes, carrots and beans, and maintains a database that stores the names of the vegetables bought in each customer purchase, as shown in *Table 1*. For this scenario, the itemset I corresponds to the set of vegetables that are offered for sale. Then, on mining the database we will find rules of the following nature:

$$Potatoes, Onions \Rightarrow Tomatoes \mid (75, 60)$$

$$Beans \Rightarrow Potatoes \mid (80, 40)$$

which translate to “Seventy-five percent of customers who bought potatoes and onions also bought tomatoes. Sixty percent of the customers made such purchases” and “Eighty



Box 3 Large Database Mining Example*

A group of researchers recently collected sales data from a large retailing company located in the U.S.A. There were 46,873 customer transactions in this data. Each transaction contained the names of the departments from which a customer bought an item during a visit. The store comprised of 63 departments. Rule mining was executed on this data to determine the associations between departments in the customer purchasing behavior.

After mining, the following rules were found for a minimum support of 1 percent and minimum confidence of 50 percent.

Tires	⇒	Automotive Services	(98.80, 5.79)
Auto Accessories, Tires	⇒	Automotive Services	(98.29, 1.47)
Auto Accessories	⇒	Automotive Services	(79.51, 11.81)
Automotive Services	⇒	Auto Accessories	(71.60, 11.81)
Home Laundry Appliances	⇒	Maintenance Agreement Sales	(66.55, 1.25)
Children's Hardlines	⇒	Infant and Children's Wear	(66.15, 4.24)
Men's Furnishings	⇒	Men's Sportswear	(54.86, 5.21)

*This example is taken from Agrawal et al (May 1993) found in the Suggested Reading list.

percent of the customers who bought beans also bought potatoes. Forty percent of the customers made such purchases", respectively.

A word of caution: The rules derived in the above example are not truly valid since the database used in the example is a *toy* database that has only ten customer records – it was provided only for illustrative purposes. For rules to be meaningful, they should be derived from large databases that have thousands of customer records, thereby indicating consistent patterns, not transient phenomena. A real-world example of rules mined from a large database is shown in *Box 3*.

Rule Discovery

As mentioned earlier, identifying rules based on patterns embedded in the historical data can serve to improve business



For rules to be meaningful, they should be derived from large databases that have thousands of customer records, thereby indicating consistent patterns, not transient phenomena.

decisions. Of course, rules may sometimes be obvious or common-sense, that is, they would be known without mining the database. For example, the fact that butter is usually bought with bread is known to every shopkeeper. In large organizations, however, rules may be more subtle and are perceived only after mining the database.

Given the need for mining historical databases, we would obviously like to implement this in as efficient a manner as possible since searching for patterns can be computationally very expensive. Therefore, the main focus in data mining research has been on designing efficient rule discovery algorithms.

The inputs to the rule discovery problem are I , a set of items, and D , a database that stores transactional information about these items. In addition, the user provides the values for sup_{min} , the minimum level of support that a rule must have to be considered significant by the user, and con_{min} , the minimum level of confidence that a rule must have in order to be useful. Within this framework, the rule mining problem can be decomposed into two sub-problems:

Frequent Itemset Generation

Find all combinations of items that have a support factor of at least sup_{min} . These combinations are called *frequent itemsets*, whereas all other combinations are called *rare itemsets* (since they occur too infrequently to be of interest to the user).

Strong Rule Derivation

Use the frequent itemsets to generate rules that have the requisite strength, that is, their confidence factor is at least con_{min} .

In the following two sections, techniques for solving each of the above sub-problems are presented.

Frequent Itemset Generation

A simple and straightforward method for generating frequent itemsets is to make a *single pass* through the entire database, and in the process measure the support for every itemset in the database. Implementing this solution requires the setting up of a measurement counter for each subset of the set of items I that occurs in the database, and in the worst case, when every subset is represented, the total number of counters required is 2^M , where M is the number of items in I . Since M is typically of the order of a few hundreds or thousands, the number of counters required far exceeds the capabilities of present-day computing systems. Therefore, the *one-pass* solution is clearly infeasible, and several *multi-pass* solutions have therefore been developed.

A simple multi-pass solution works as follows: The algorithm makes multiple passes over the database and in each pass, the support for only certain specific itemsets is measured. These itemsets are called *candidate itemsets*. At the end of a pass, the support for each of the candidate itemsets associated with that pass is evaluated and compared with sup_{min} (the minimum support) to determine whether the itemset is frequent or rare.

Candidate itemsets are identified using the following scheme: Assume that the set of frequent itemsets found at the end of the k th pass is F_k . Then, in the next pass, the candidate itemsets are comprised of all itemsets that are constructed as *one-extensions* of itemsets present in F_k . A one-extension of an itemset is the itemset extended by exactly one item. For example, given a set of items A, B, C, D and E , the one-extensions of the itemset AB are ABC , ABD and ABE . While this scheme of generating candidate itemsets works in general, in order to start off the process we need to prespecify the candidate itemsets for the very first pass ($k = 1$). This is done by making each individual item in I a candidate itemset for the first pass.

Since searching for patterns can be computationally very expensive, the main focus in data mining research has been on designing efficient rule discovery algorithms.



The basic idea in the above procedure is simply that ‘If a particular itemset is found to be rare, then all its extensions are also guaranteed to be rare’. This is because the support for an extension of an itemset cannot be more than the support for the itemset itself. So, if AB is found to be rare, there is no need to measure the support for ABC , ABD , $ABCD$, etc., since they are also certain to be rare. Therefore, in each pass, the search space is *pruned* to measure only those itemsets that are potentially frequent and the rare itemsets are not considered further.

Another feature of the above procedure is that in the k th pass over the database, only itemsets that contain exactly k items are measured, due to the one-extension approach. This means that no more than M passes are required to identify all the frequent itemsets resident in the historical database.

Algorithms that are even better than the one outlined above have been proposed in recent database conferences. In fact, one paper presents a novel technique by which all frequent itemsets are generated in just two passes! Moreover, different approaches to database mining have also been investigated (one of these is described in *Box 4*).

Strong Rule Derivation

In the previous section, we described methods for generating frequent itemsets. We now move on to the second sub-problem, namely that of deriving strong rules from the frequent itemsets. The rule derivation problem can be solved using the following simple method: For every frequent itemset F , enumerate all the subsets of F . For every such subset f , output a rule of the form $f \Rightarrow (F-f)$ if the rule is sufficiently strong. The strength is easily determined by computing the ratio of the support factor of F to that of the support factor of f . If this value is at least con_{min} , the minimum rule confidence factor, the rule is considered to be strong and is displayed to the user, otherwise it is discarded.

Box 4 Database Sampling

At first glance, it would appear that rules of the type discussed in this article can be derived easily by using well-known statistical methods, for example, sampling. In sampling, inferences about an entire population are made based on characteristics exhibited by a representative subset of the population. This approach is especially attractive for data mining since, instead of scanning the whole database, only a small part of it has to be processed, thereby leading to much better performance.

The main drawback of the sampling approach is that it may not identify *all* the applicable rules, especially when the minimum support factor specified by the user is low. So, for example, if the minimum support factor is one percent, then sampling will usually identify all the high support rules but miss out on rules whose support is low (below five percent, say). This is because the chosen subset of the database may not reflect these low support rules. Recent research indicates that for a minimum support of one percent, using even a sample as large as 50 percent of the database is not sufficient to ensure that all rules are generated !

If complete accuracy is not required by the user, however, then database sampling can be fruitfully incorporated into the data mining framework by executing the data mining algorithm on the sample rather than on the original database.

In the above procedure, the only part that is potentially difficult is the enumeration of all the subsets of each frequent itemset. However, efficient algorithms are available for performing this task and therefore the rule derivation problem is easy to handle.

Classification and Sequence Rules

The rules that we have discussed so far are called *association rules* since they involve finding associations between sets of items. However, they are only one example of the types of rules that may be of interest to an organization. Other interesting rule classes that have been identified in the literature are *classification rules* and *sequence rules*, and data mining algorithms for discovering these types of rules have also been developed in the last few years.



The goal of Database Mining is to discover information from historical organizational databases that can be used to improve their business decisions.

The classification problem involves finding rules that *partition* the data into disjoint groups. For example, the courses offered by a college may be categorized into good, average and bad based on the number of students that attend each course. Assume that the attendance in a course is primarily based on the qualities of the teacher. Also assume that the college has maintained a database about the attributes of all its teachers. With this data and the course attendance information, a profile of the attributes of successful teachers can be developed. Then, this profile can be used by the college for short-listing the set of good candidate teachers whenever new courses are to be offered. For example, the rule could be *if a candidate has a master's degree, is less than 40 years of age, and has more than 5 years experience, then the candidate is expected to be a good teacher.*

Organizations quite often have to deal with *ordered* data, that is data that is sorted on some dimension, usually time. Currency exchange rates and stock share prices are examples of this kind of data. Rules derived from ordered data are called sequence rules. An example is *when the value of the US dollar goes up on two consecutive days and the British pound remains stable during this period, the Indian rupee goes up the next day 75 percent of the time.*

Summary

The goal of Database Mining is to discover information from historical organizational databases that can be used to improve their business decisions. Developing efficient algorithms for mining has become an active area of research in the database community in the last few years. Although commercial data mining packages are not yet available, there are several research prototypes that have been developed. Examples are QUEST constructed at IBM's Almaden Research Center in San Jose, California, U.S.A., and DISCOVER, available from the Hong Kong University of Science and Technology. We expect that

sophisticated database mining packages will be available soon and that they will become essential software for all organizations within a few years.

Acknowledgements

The material presented in this paper is mainly derived from the publications mentioned in the Suggested Reading list. This work was supported in part by a research grant from the Department of Science and Technology, Govt. of India.

Suggested Reading

- ◆ H Korth and A Silberschatz. *Database System Concepts*, 2nd ed., McGraw-Hill, 1991.
- ◆ V Rajaraman. *Analysis and Design of Information Systems*. Prentice-Hall India, 1991.
- ◆ R Agrawal, T Imielinski and A Swami. Mining Association Rules between Sets of Items in Large Databases. *Proceedings of 22nd ACM SIGMOD International Conference on Management of Data*, Washington D.C., U.S.A., May 1993.
- ◆ R Agrawal, T Imielinski and A Swami. Database Mining: A Performance Perspective. *IEEE Transactions on Knowledge and Data Engineering*, December 1993.
- ◆ J Gray and A Reuter. *Transaction Processing: Concepts and Techniques*. Morgan Kaufmann, 1993.
- ◆ R Agrawal and R Srikant. Fast Algorithms for Mining Association Rules. *Proceedings of 20th Very Large Data Base Conference*, Santiago, Chile, September 1994.
- ◆ R Elmasri and S Navathe. *Fundamentals of Database Systems*, 2nd ed., Addison-Wesley, 1994.
- ◆ A Savasere, E Omiecinski and S Navathe. An Efficient Algorithm for Mining Association Rules in Large Databases. *Proceedings of 21st Very Large Data Base Conference*, Zurich, Switzerland, September 1995.

Address for Correspondence
 Jayant R Haritsa
 Supercomputer Education
 and Research Centre
 Indian Institute of Science
 Bangalore 560 012, India

Classroom



In this section of Resonance, we invite readers to pose questions likely to be raised in a classroom situation. We may suggest strategies for dealing with them, or invite responses, or both. "Classroom" is equally a forum for raising broader issues and sharing personal experiences and viewpoints on matters related to teaching and learning science.

! Quantum Theory of the Doppler Effect

GS Ranganath, Raman Research Institute, Bangalore 560 080

Generally text books give only the wave theory of the Doppler effect. It is instructive to consider the same phenomenon in the quantum theory, allowing for the effects of special relativity. Let M be the mass of the source and V its velocity before photon emission. When a photon is emitted by the source, the internal energy E changes to E' . We define $\nu_0 = \frac{E-E'}{h}$. As a result of photon emission the source experiences a recoil, and hence its velocity changes to V' (Figure 1). In the relativistic case the change in the internal energy of the source, is nothing but the energy associated with the change in the rest mass of the source. If the rest masses of the source before and after photon emission be M and M' respectively, then $(E - E') = (M - M') c^2$.

Now momentum conservation leads to

$$\frac{MV}{\sqrt{1-\beta^2}} = \frac{M'V'}{\sqrt{1-\beta'^2}} \cos \phi + \frac{h\nu}{c} \cos \Theta,$$

$$0 = \frac{M'V'}{\sqrt{1-\beta'^2}} \sin \phi - \frac{h\nu}{c} \sin \Theta,$$

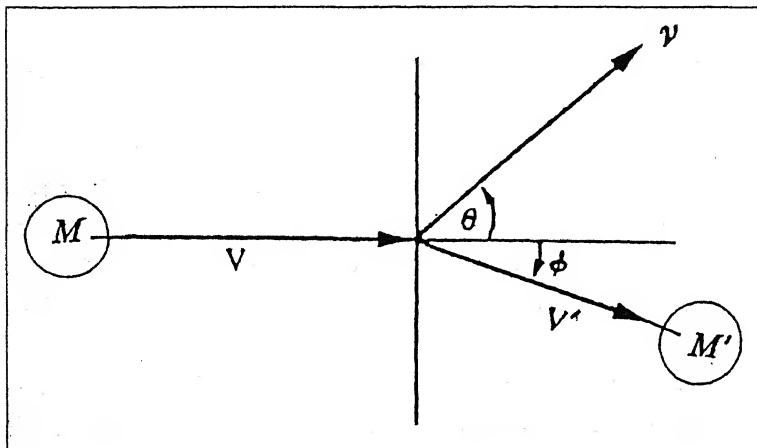


Figure 1 Quantum theory of the Doppler effect.

Further, energy conservation gives

$$\frac{Mc^2}{\sqrt{1-\beta^2}} = \frac{M'c^2}{\sqrt{1-\beta'^2}} + h\nu.$$

In these equations $\beta = V/c$, $\beta' = V'/c$ and c is the velocity of light. Also Θ is the angle between the direction of the initial velocity V and the direction of the light emission and ϕ is the angle between the directions of V and V' . Eliminating M' , V' and ϕ , we get

$$\frac{Mc^2\nu}{\sqrt{1-\beta^2}}(1 - \beta \cos \Theta) = Mc^2\nu_0 - \frac{h\nu_0^2}{2}.$$

If the mass M of the source is sufficiently large compared to the photonic mass ($h\nu_0/c^2$) of light then we can neglect the last term to get:

$$\nu = \nu_0 \frac{\sqrt{1-\beta^2}}{1 - \beta \cos \Theta}.$$

It is important to note that $\nu \neq \nu_0$ even when $\Theta = \pi/2$. This is the famous relativistic transverse Doppler effect. Incidentally we get the well known result of the non-relativistic Doppler effect when

$\Theta = 0$ and β^2 is negligible compared to unity. This theory is rather reminiscent of the Compton theory of X-ray scattering by electrons.

We can easily generalize this theory to discuss the Doppler effect in a medium. If the refractive index of the medium is μ then we change β to $\mu\beta$. An interesting possibility in a medium is that the source can move with a velocity greater than the velocity of light in that medium i. e., $V > c/\mu$. Then a careful analysis leads to the surprising result that even if the source is not initially excited, it will be excited with the simultaneous emission of photons inside the Cherenkov cone ($\Theta < \Theta_0 = \cos^{-1}(\frac{1}{\mu\beta})$). Then the excited source will make a transition to the lower state emitting photons outside the Cherenkov cone ($\Theta > \Theta_0$). This is not a paradoxical result since the energy needed for exciting the source and for the emission of the photon are both derived from the kinetic energy of the source.

Suggested Reading

- ◆ V L Ginzburg, L M Levin, M S Rabinovich and D V Sivukhin. *Problems in undergraduate physics*. Pergamon Press, 1965.
- ◆ V L Ginzburg. *Waynflete Lectures on Physics*. Pergamon Press, 1983.

! Microbiology as if Bird Watching

Milind G Watve, M.E. Society,
Abasaheb Garware College,
Karve Road, Pune 411 004, India.

I became a bird-watcher much before I started studying Microbiology for my Bachelor's. I wasn't sure why I opted for Microbiology. Perhaps just by default, since in those days, there were few options for a student not going into medicine. I was equally uncertain about taking bird-watching seriously. It just so happened that one of our instructors, in a sort of hobby-club to which I subscribed in my high school days, took us bird watching once or twice. Somehow, this hobby lingered on and became more and more absorbing. Today I can say with confidence that if I am a good microbiologist, it is because of bird-watching.

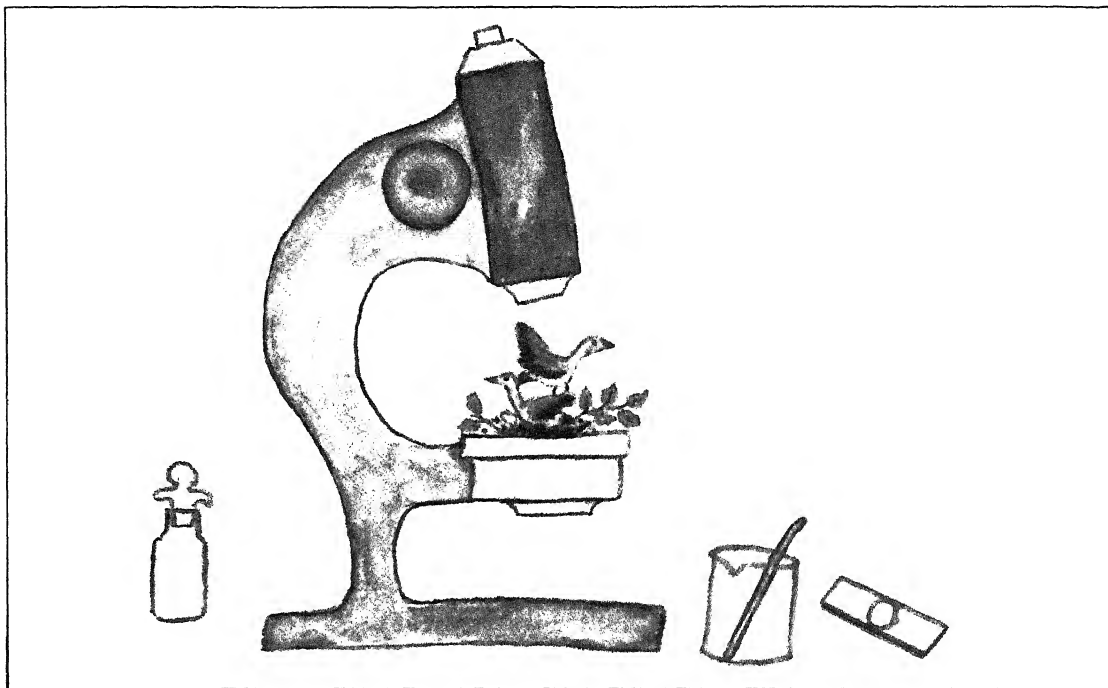
When we started our course in Microbiology, I found it extremely monotonous, dry and rigid. I always compared it with bird-watching which was juicy, enlightening and ever-challenging. I always wondered what caused the difference. It might have been the beauty and the grace of birds, or the fact that they had simple names in English and vernacular so the Latin names did not trouble us, or just the feeling of being out of doors, or was it something else?

Today I think I have identified the most important reason. Our microbiology teacher showed us — “This is *E. coli*. It forms pink colonies on MacConkey’s agar because it ferments lactose and produces acid. *Salmonella* produces colourless colonies because it does not ferment lactose.” The bird watching instructor, on the other hand, had a different approach. “What will you do if you see a strange bird? Note its size in comparison with known birds. Look at any peculiar features — the beak, the tail, shape of wings, style of flight, whether it hops or walks when on the ground. Draw and describe whatever you see. Then go to the library, get Salim Ali’s or any other bird book and discover yourself.” This I think was the key. Our formal training in Microbiology, Zoology or Botany seldom encouraged us to ‘discover things by ourselves’, be it a different type of colony, an odd insect or a weird mushroom.

Fortunately our M.Sc. course was full of opportunities for exploration. Ours was the first batch in the post graduate course at our centre. Everything was crude, unstandardized and uncertain. This, I think, was the best that could have happened to us. Since we had to take the initiative, we started exploring. We very soon came to realise that the microbial world was in no way less fascinating than birds. In fact it was much more diverse, beautiful and lively. There weren’t only rods and cocci, there were the charming rosettes of stalked bacteria, the beaded spirals of actinomycete spores, the army ant like scavenging swarms of gliding bacteria, the never ending long beeggiatoa, the star-shaped, pear-shaped, needle-shaped and ribbon-shaped odd

Since our M.Sc course was the first one everything was crude, unstandardized and uncertain. This, I think, was the best that could have happened to us. Since we had to take the initiative, we started exploring.





species and of all the brightest colourful flowers of the earth—the myxobacterial fruiting bodies! And all of them were very easy to find. The key to finding them was to be always curious and be on the look out and venture beyond the text-book methods.

If exploration is a necessary element of science and having a standardized curriculum is detrimental to exploration, perhaps having a curriculum itself is against the spirit of science.

Exploration led to frustration at times when the forms observed did not match anything in the literature. Several years later I read that above 90 % of the microbial world remains unexplored. There are 4000 different types of DNAs (and therefore as many types of microbes) to be found in one gram of soil. If we plate out, we get some 50–60 types of colonies. Many of them are unknown to science. Stretch your imagination, try a novel method of cultivation and you are almost certain to hit upon a cluster of strange species.

Unfortunately our Microbiology training seldom goes beyond *E. coli*. But this is nothing peculiar to Microbiology. While the spirit of science has to begin with exploration, the component of

RESONANCE

Questionnaire for Readers

1. Your age: < 15 15-20 20-25 25-30 30-40 40-50 50-60 >60

2. To which category do you belong?

a) Student in high school XI/ XII/PUC/plus 2 BSc MSc

Engineering/Medicine PhD Other

b) Teacher in VIII-X XI-XII/PUC/plus 2 UG PG Other

c) Scientist in Univ./Research Institute, R & D/Industry Other

d) Other (please specify)

3. Your subjects:

1. Biology

4. Mathematics

2. Chemistry

5. Physics

3. Engineering/Medicine

6. Other

4. Do you subscribe to *Resonance*? Yes/No

5. How many issues of *Resonance* have you read so far?

1 2 3 4 5 6 7 8

6. Tell us what you read in *Resonance*

a) Biology

e) Physics

b) Chemistry

f) Other

c) Computer Science & Engineering

g) Everything

d) Mathematics

7. In what ways has *Resonance* helped you?

a) In understanding the subject

c) In increasing general knowledge

b) In your studies

d) Anything else (please specify)

8. What do you think the length of an article (no. of pages) ought to be?

9. How would you rate the general quality of articles in each subject (in a scale of 1 to 5, 1 being poor, 5 being outstanding)? Circle your choice in each case

a) Biology

1 2 3 4 5

d) Mathematics

1 2 3 4 5

b) Chemistry

1 2 3 4 5

e) Physics

1 2 3 4 5

c) Computer Sci. & Engg.

1 2 3 4 5

f) Classroom

1 2 3 4 5



10. In your opinion which categories of students are able to read and understand at least 50% of the articles?

- a) Plus two b) Undergraduate c) PG d) PhD

11. How much of the published material is directly usable in the classroom?

- <25% 25-50% >50%

12. What would you like to see more of in *Resonance*? (Tick as many as applicable)

- | | | |
|---------------------|------------------|----------------|
| a) Series articles | e) Experiments | i) Reflections |
| b) General articles | f) Book reviews | j) Any other |
| c) Features | g) Classroom | |
| d) Research News | h) Think it Over | |

13. At present the subscription to *Resonance* is highly subsidized. This may not be possible for a long period. How much do you think would be a fair subscription for 12 monthly issues of *Resonance*?

Individual, Rs.

Institutional, Rs.

14. Would you like *Resonance* to appear

- a) Monthly as now b) Once in two months

15. Any other comments:

(please be brief; use additional sheets if required.)

Name: _____

Address: _____

This form may please be completed and returned to:

The Chief Editor
Resonance
Indian Academy of Sciences
Post Box No. 8005
C V Raman Avenue
Bangalore 560 080, India



exploration is missing in our science training. Part of the problem is that the present university structure demands that there should be uniformity in teaching in all the colleges in one university. It is therefore necessary that the syllabus specifies every detail to be taught. It follows naturally, or at least as the interpretation of students and teachers goes, that other things are not to be taught. In effect the stuff to be studied is precisely defined, rigidly standardized and that marks the end of exploration! You have only so and so species, such and such experiment in the curriculum and that is the final word. It follows logically that if exploration is a necessary element of science and having a standardized curriculum is detrimental to exploration, perhaps having a curriculum itself is against the spirit of science.

Can we teach science without a curriculum, or at least without a rigid curriculum? I tried it once when my boss asked me to conduct first year practicals. On the first day I handed over a microscope and a screw-driver to each group and asked them to play with it. Dismantle every part possible, and try to put everything back. On the next day I asked them to bring anything they would like to see under a microscope and observe it. It could be any crazy object. Thereafter things took their own shape. We observed a large variety of microbes. Obviously, the question that followed was "what is this?". I did not answer the question. In fact, many times I could not. The organisms were so diverse and different that I had to say — I don't know. Probably nobody knows. Imagine that you are the first biologist on earth to see such a creature. How will you describe it? How will you draw it? That's almost like a bird-watcher looking into a microscope. The students were quite fascinated by what they were doing. From observations came questions. The questions prompted more observations and experiments. We did whatever everybody thought was the next logical step. The students did not prepare Nutrient Agar or MacConkey's agar, they prepared whatever they thought suited their organism. At the end of the year we looked at the syllabus and found that it had been already covered. Microbiology can be as exciting as bird watching!

Can we teach science without a curriculum, or at least without a rigid curriculum?



Think It Over



This section of Resonance is meant to raise thought-provoking, interesting, or just plain brain-teasing questions every month, and discuss answers a few months later. Readers are welcome to send in suggestions for such questions, solutions to questions already posed, comments on the solutions discussed in the journal, etc. to Resonance Indian Academy of Sciences, Bangalore 560 080, with "Think It Over" written on the cover or card to help us sort the correspondence. Due to limitations of space, it may not be possible to use all the material received. However, the coordinators of this section (currently A Sitaram and R Nityananda) will try and select items which best illustrate various ideas and concepts, for inclusion in this section.

J Chandrasekhar, (born on October 23) Department of Organic Chemistry, Indian Institute of Science, Bangalore.

Counting Molecules in a Spoonful of Water

What do 18 g of water, 27 g of aluminium, 197 g of gold and 342 g of sugar have in common? The quantities correspond to the atomic weight of the metals and the molecular weight of the molecular substances in grams. These measures represent one mole of each material. They all contain the same number of atoms or molecules. To be precise, 602,205,000,000,000,000,000 of them, give or take a few trillion-millions. This astronomical figure is the famous *Avogadro Number*.

The number appears in text books during the discussion of gas laws. It is the number of molecules in 22.4 litres of a gas at 25° C and 1 atmosphere pressure. As mentioned earlier, it is also the number of molecules in a gram molecular weight of any substance. The magical figure shows up in electrochemistry too. An Avogadro number of electric charge (called a Faraday) is needed to deposit a mole of a metal from a solution containing its monocation. Thus,

6.02×10^{20} electrons would be needed to electroplate a surface with 0.197 g of gold.

The concept of mole is very important in chemistry. Atoms and molecules usually combine in simple molar ratios. Therefore, the optimum way to understand chemical reactions and equilibria is by determining the number of moles of each of the species involved. The proportions in terms of weight or volume are more complex, but can be calculated from the molar ratios.

In recent times, October twenty third (written 10/23 by Americans) of each year is celebrated as the International Mole Day (no, not the Zoology variety). The festivities begin at two minutes past six (AM or PM, as per convenience). This specific time in a thoroughly unimportant date in history has been chosen to create greater awareness for arguably the most important number in chemistry: 6.02×10^{23} .

It is all very well, but how can the value of Avogadro number be determined experimentally?



Smallest Possible Magic Square of Consecutive Odd Prime Numbers

1	823	821	809	811	797	19	29	313	31	23	37
89	83	211	79	641	631	619	709	617	53	43	739
97	227	103	107	193	557	719	727	607	139	757	281
223	653	499	197	109	113	563	479	173	761	587	157
367	379	521	383	241	467	257	263	269	167	601	599
349	359	353	647	389	331	317	311	409	307	293	449
503	523	233	337	547	397	421	17	401	271	431	433
229	491	373	487	461	251	443	463	137	439	457	283
509	199	73	541	347	191	181	569	577	571	163	593
661	101	643	239	691	701	127	131	179	613	277	151
659	673	677	683	71	67	61	47	59	743	733	41
827	3	7	5	13	11	787	769	773	419	149	751

In a magic square the sum of all the elements of a row or a column or a diagonal will be the same. In the present case this is 4514.

How to Move in a Jostling Crowd

The Art of Harnessing Random Motions

G S Ranganath

For people living in big cities it is an ordeal to walk in a bus stand or a railway station. They get stuck helplessly in a crowd. They are simply pushed around and all their efforts to move forward appear futile. Only the most energetic can wade through the constantly moving sea of people. How about the weaker ones? Now there is a way out even for them, provided they emulate the bug *Listeria manocyto-gen*.

These bacteria cause the disease meningitis. They have the special knack of swimming in water by exploiting the random motions of the water molecules i.e., Brownian motion. When Brownian motion kicks this bacterium forward, the microbe for a moment allows itself to be pushed. Then it quickly anchors itself in the new place until it gets another push forward. The bacterium accomplishes this extraordinary feat with the help of its bushy tail. So next time you get mixed up in a crowd you can navigate across like this deadly pathogen. Allow the crowd to push you if it is in the right direction. Otherwise stay firmly wherever you are. Thus you can skillfully get out of the mess. *Listeria's* talent for turning random thermal motions into net movement has attracted the attention of

scientists because it extracts work out of something generally regarded as useless 'noise'.

Are there other schemes for harnessing Brownian motion? In this context it is instructive to remember what Feynman said about thirty years ago. He used a ratchet as an example to argue that useful work cannot be extracted from equilibrium fluctuations. A ratchet is a mechanical device that can turn only one way. In recent years it has been realised that the same restrictions do not apply to non-equilibrium noise. In other words an asymmetric potential can rectify symmetric fluctuations. In common parlance this means that structures with spatial asymmetry can act as 'ratchets' for randomly moving particles. This means that under the influence of such a structure the randomly moving particles start drifting in a particular direction.

In view of such implications, ratchets have become important and fashionable systems to study. In 1994 Rousselet and co-workers came up with an electrical device that acted as a ratchet. They built a microelectrode

Listeria's talent for turning random thermal motions into net movement has attracted the attention of scientists because it extracts work out of something generally regarded as useless 'noise'.

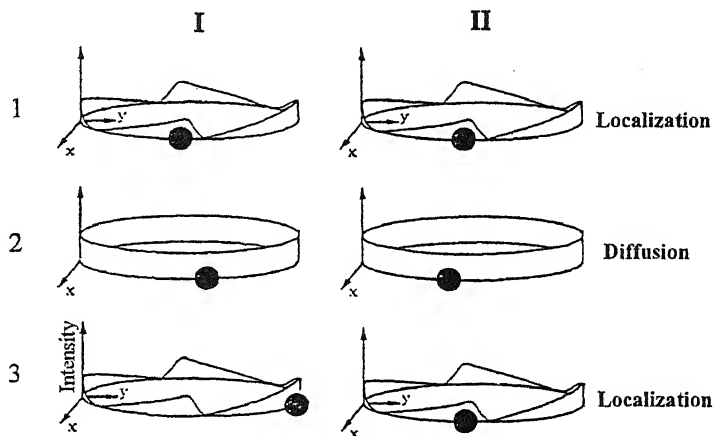
system which generated a periodically appearing and disappearing saw-tooth electrostatic potential. In a medium of randomly moving colloidal particles this device pushed forward small particles of sizes 0.1 to 0.5 microns. Interestingly, the velocity of the forward motion appeared to depend on the particle size. Hence these authors also suggested that their device can even separate particles of different dimensions. How this actually works can be best understood by considering the latest in this 'ratchet game' viz., the *Optical thermal ratchet*. This is probably the most elegant of such devices produced so far. This system was developed last year by Faucheux and his collaborators. In this experiment a plastic sphere of about 1.5 micron diameter, immersed in water, is illuminated by a circularly polarised infrared laser beam. The laser beam rapidly traces a circle in space. The electric field of the laser beam induces at every instant an electric dipole on the plastic sphere. This dipole in turn is attracted to the region of high electric field i.e. light intensity. It may be recalled that the very same mechanism is responsible for the attraction of paper bits by a comb, soon after brushing hairs. Thus the sphere will always be driven to the region of high optical intensity. In the process, the sphere will be confined to remain in the circle. However it need not stay at one place on this circle but can thermally diffuse from one point on it to another. If the incident beam is modulated by a chopper so that the light intensity gets modulated as a saw tooth wave, then all round the circle the intensity

It is important to mention here that very similar ratchet mechanisms have been suggested in the case of molecular motors like kinesins used in the transport of vesicles in cells.

varies like a saw-tooth wave. The fact that the light intensity is modulated will force the sphere to always move towards the most intensely illuminated region of the circle. That is, it will move towards the position of peak intensity and get trapped there. When the chopper is off i.e., modulation disappears, the sphere can randomly diffuse away from the position of the peak intensity where it had been trapped. Let us say that the sphere diffuses past what was earlier a steep drop-off. Then it will move forward up the gentle slope to the next peak when the chopper is turned on. Instead, let us say it diffuses in the opposite direction. Then it will return to its original position when the chopper is on. Now the light modulation is such that every saw tooth is turned the same way. Thus, when the modulated light is continuously turned on and off rapidly, we periodically create, a saw-tooth modulation in an otherwise uniformly bright circle. Hence the particle can be made to migrate in one direction only, up the gentle slope. It is important to mention here that very similar ratchet mechanisms have been suggested in the case of molecular motors like kinesins used in the transport of vesicles in cells.

Optical Ratchet

- 1 A plastic sphere in an optical trap at the peak intensity of a saw-tooth profile.
- 2 When modulations are absent we get random drift, due to Brownian motion, to the right (I) or left (II).
- 3 Reappearance of the saw-tooth profile either pushes the sphere forward (I) or returns it to the same peak (II).



We can suggest in a lighter vein (with apologies to mathematicians) an interesting solution to the problem of a drunkard's walk in one dimension (See *Resonance* July 1996). Though the drunkard will reach his house eventually, he will take an unduly long time. We can speed up his homecoming. Let us say that the path from the bar to the drunkard's house is paved with identical planks all of them hinged at the ends facing the house. The other end of each plank which is facing the bar can be raised to a pre-determined height. Therefore, when all the planks are raised at the ends facing the bar, we find the path to be in the nature of periodic gentle slopes like the rooftop of a workshop. Periodic raising and lowering of these planks while

the drunkard is walking is exactly like the previously discussed case. Except that, here he gets trapped in a valley instead of a peak. Therefore by periodic raising and lowering of the planks we will be forcing the drunkard to go in the direction of his house!

Suggested Reading

- ◆ R P Feynman, R B Leighton and M Sands. *Feynman Lectures on Physics*, Vol. I. Chapter 6. Addison Wesley, 1963.
- ◆ J Rousselet et al. *Nature*. Vol. 370, p. 446, 1994.
- ◆ L P Faucheux et al. *Phys. Rev. Letts*. Vol. 74, p. 150, 1995.

G S Ranganath is with Raman Research Institute, Bangalore 560 080.

Harmonic Analysis

Fourier Series and Beyond

KR Parthasarathy



Harmonic Analysis

Henry Helson

No.7, Texts and Readings in
Mathematics,

Hindustan Book Agency,
New Delhi 110 007, 1995.

Second edition. pp.227, Rs.235

While investigating the properties of heat flow in 1807 the French scientist Jean Baptiste Joseph Fourier stumbled on the remarkably fruitful mathematical idea that the graph of any function in a bounded interval can be obtained as a linear superposition of sines and cosines. Since $\cos x + i \sin x = e^{ix}$ this led to the hypothesis that any integrable function f in the interval $[0, 2\pi]$ can be expanded as a Fourier series:

$$f(x) = \sum_n a_n e^{inx}, \quad (1)$$

where $a_n = a_n(f)$ is the n th Fourier coefficient defined by

$$a_n = (2\pi)^{-1} \int_0^{2\pi} f(x) e^{-inx} dx,$$

Reprinted from Current Science

While investigating the properties of heat flow in 1807 the French scientist Jean Baptiste Joseph Fourier stumbled on the remarkably fruitful mathematical idea that the graph of any function in a bounded interval can be obtained as a linear superposition of sines and cosines.

where

$$n = 0, \pm 1, \pm 2, \dots \quad (2)$$

An infinite series can have several interpretations depending on the choice of the notion of its convergence. The investigation of convergence properties of the Fourier series (1) led to a vast amount of mathematical literature including the theory of the Lebesgue integral. The first chapter of Helson's little volume on Harmonic analysis provides a quick survey of these developments including the classical kernels of Dirichlet, Fejèr and Poisson, the general notion of an approximate identity in a convolution algebra and a result of the author and A Beurling on measures with bounded powers.

It is important to note that the set \mathbb{Z} of all integers is a discrete abelian (or commutative) group under addition and

discrete topology whereas the set T of all complex numbers of modulus unity is a compact abelian group under multiplication and the relative topology inherited from the complex plane. Furthermore, the function $B(n, z) = z^n$ on $\mathbf{Z} \times T$ has the properties

$$B(m+n, z) = B(m, z)B(n, z),$$

$$B(n, z_1 z_2) = B(n, z_1)B(n, z_2).$$

The map $z \rightarrow B(n, z)$ is a continuous homomorphism from T into itself for each n and every continuous homomorphism of T into itself is accounted for in this list. Similarly, the map $n \rightarrow z^n$ is a (continuous) homomorphism from \mathbf{Z} into T and every homomorphism from \mathbf{Z} into T is of this kind. The Fourier series (1) can now be expressed as

$$f(z) = \sum_{n \in \mathbf{Z}} a_n B(n, z), \quad z = e^{iz} \in T \quad (3)$$

after identifying T with the interval $[0, 2\pi]$, where 0 and 2π on the line represent the same point 1 on T . Expressed in this way one has the following generalization of (3): Suppose G is any abelian group with the group operation denoted by $+$. Then there exists a discrete abelian group \hat{G} whose operation is viewed as multiplication and a map $B(., .)$ from $\hat{G} \times G$ into T such that

$$B(\chi_1 \chi_2, x) = B(\chi_1, x)B(\chi_2, x),$$

$$B(\chi, x_1 + x_2) = B(\chi, x_1)B(\chi, x_2).$$

for all $\chi, \chi_1, \chi_2 \in \hat{G}$ and $x, x_1, x_2 \in G$. The group G admits a unique (group translation invariant probability measure σ (called the Haar measure of G) and any σ -square integrable function on G admits a Fourier-like expansion

$$f(x) = \sum_{\chi \in \hat{G}} a(\chi) B(\chi, x), \quad (4)$$

where $a(\chi)$ is the Fourier coefficient of f given by

$$a(\chi) = \int_G \overline{B(\chi, x)} f(x) d\sigma(x) \quad (5)$$

and convergence of (4) is in the mean square sense. The classical Fourier series (1) is a special case of (4) where $G = T, \hat{G} = \mathbf{Z}$. However, there are no obvious analogues of the Dirichlet, Fejèr and Poisson kernels here owing to the lack of order in \hat{G} . \hat{G} is called the *dual group* of G or the group of *characters* of G . Chapter 3 of this volume presents a very readable survey of this generalization which is easily accessible for our MSc students and college teachers who have the required curiosity to explore the possibilities outside their customary examination-oriented syllabi. As applications Helson provides new proofs of three old theorems: (i) Kolmogorov's extension theorem for consistent family of finite dimension probability measures;

Chapter 3 of this volume presents a very readable survey which is easily accessible for our MSc students and college teachers who have the required curiosity to explore the possibilities outside their customary examination-oriented syllabi.

(2) Banach–Steinhaus’ uniform boundedness principle for a sequence of linear operators; (3) Minkowski’s theorem that any convex body in \mathbf{R}^n , which is symmetric about the origin and has volume $> 2^n$, has a lattice point other than the origin (the proof being due to C L Siegel and based on trigonometric sums).

It is to be noted that there exists a far reaching group-theoretic generalization of the expansion (4) when G is an arbitrary compact (but not necessarily abelian) group. This is known as the Peter–Weyl theory of which a glimpse of the abstract side is provided in the book *A Course on Topological Groups* by K Chandrasekharan which has recently appeared as TRIM 9 in the same series as the present volume. For the practical and computational aspects of this theory my favourite volume is *Group Theory and Physics* by S Sternberg (Cambridge University, Paperback Edition, 1995). Thanks to the contributions of Weyl, Wigner, Bargmann, Harish-Chandra, Gelfand and several

other mathematicians and physicists, group-theoretic harmonic analysis is a flourishing industry today paving the way to new developments in the context of noncompact Lie groups as well as quantum groups.

Since $B(n, z) = z^n$ the expansion (3) suggests a link between Fourier series and the theory of analytic functions of a complex variable. This leads to the notion of the Hardy spaces $H^p(T)$, $1 \leq p < \infty$. $H^p(T) \subset L^p(T)$ is the subspace consisting of functions f for which the Fourier coefficients a_n in (2) vanish for $n < 0$. If $B(1, z) = \chi(z)$, $z \in T$, then for any $f \in L^2(T)$ denote by M_f the closed linear span of $\{f, \chi f, \chi^2 f, \dots\}$. Then $M_f \subset L^2(T)$ is invariant under multiplication by χ . If $M_f = L^2(T)$ then f is called an *outer function*. It is a theorem of Beurling that a subspace $M \subset L^2(T)$ invariant under multiplication by χ belongs to one of two types. Either it consists of all functions in $L^2(T)$ with support in a fixed measurable subset of T or it is $qH^2(T)$ for some function q of modulus unity. The first kind of subspace is called a *Wiener subspace* and the second, a *Beurling subspace*. Any nonnull element f of $H^2(T)$ can be factorized as $f = qg$, where q and g belong to $H^2(T)$, g is outer and q is of modulus unity. Elements of $H^2(T)$ which are of unit modulus are called *inner functions*. This

factorization into an inner and an outer function is unique up to a constant factor of modulus unity. This implies that every nonnull element of $H^1(T)$ can be factorized as qg^2 , where q is inner and g is outer in $H^2(T)$. A function f in $H^2(T)$ is outer if and only if

$$\int_0^{2\pi} \log |f(e^{ix})| dx > -\infty.$$

From these results of Beurling it is possible to deduce the following theorem due to G Szegö: If w is a nonnegative integrable function on T then

$$\exp \frac{1}{2\pi} \int_0^{2\pi} \log w(e^{ix}) dx = \inf_P \frac{1}{2\pi} \int_0^{2\pi} |(1 + P(e^{ix}))|^2 w(e^{ix}) dx;$$

where P ranges over all polynomials. The densely packed chapter on Hardy spaces covers all these and much more. It may be noted that this last theorem of Szegö is at the heart of the theory of prediction of discrete time one-dimensional stationary stochastic processes developed by N Wiener in the US and A N Kolmogorov in the former USSR during the Second World war. (A multidimensional version of Szegö's theorem when w is a positive definite matrix-valued function on T with summable entries was obtained by N Wiener and P Masani when they met at the Indian Statistical Institute in

Thanks to the contributions of Weyl, Wigner, Bargmann, Harish-Chandra, Gelfand and several other mathematicians and physicists, group-theoretic harmonic analysis is a flourishing industry today paving the way to new developments in the context of noncompact Lie groups as well as quantum groups.

Calcutta during 1955–56.) By exploiting the standard conformal map from the unit disk to the upper half plane the author indicates how a theory of Hardy spaces $H^p(\mathbf{R})$ could be built. (This could be used to develop the prediction theory of one dimensional continuous time stationary stochastic processes.)

A fairly extensive discussion of the theory of conjugate functions in a whole chapter is followed by a brief account of $(\mathbf{R}$ and $\mathbf{R}_+)$ translation invariant subspaces of $L^2(\mathbf{R})$ and $L^1(\mathbf{R})$ covering the results of Wiener, Beurling and Titmarsh.

If $\varphi \in L^{\text{inf}}(\mathbf{R})$ then its Fourier transform $\hat{\varphi}$ is a tempered distribution on \mathbf{R} and the support of $\hat{\varphi}$ is called the *spectral set* of φ . The spectral set of a bounded bilateral sequence, i.e. an element of $l^\infty(\mathbf{Z})$ can be similarly defined.

With its well punctuated historical comments and instructive exercises this little but very rich volume offers an enjoyable guided tour of classical harmonic analysis

as a subset of T . An element $\varphi \in L^\infty(\mathbf{R})$ has exactly one point λ in its spectral set if and only if $\varphi(x) = \exp i\lambda x$. If $\{\alpha_n\}$ is a bilateral sequence whose terms are drawn from a finite set of complex numbers then its spectral set is the whole of T unless $\{\alpha_n\}$ is periodic. A bilateral sequence of 0's and 1's is the Fourier-Stieltjes transform of a complex measure on T if and only if it is periodic after dropping a finite number of terms. Pretty surprises of this kind are strewn around in several places in this flower garden of harmonic analysis.

Helson concludes with a little chapter on equidistribution theorems originating in the work of H Weyl. A sequence $\{u_k\}$, $k \geq 1$ in $[0, 1]$ is said to be *equidistributed* if for any interval $[a, b] \subset [0, 1]$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \# \{j \mid u_j \in [a, b], 1 \leq j \leq n\} = b - a,$$

where $\#$ denotes cardinality. To verify

the equidistribution of a real sequence $\{u_k\}$ modulo 1 it is enough to check that

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n e^{2\pi i j u_k} = 0$$

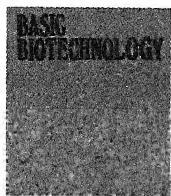
for every $j \neq 0$. Thus equidistribution and trigonometric sums are closely related. It is a theorem of van der Corput that a sequence $\{u_k\}$ is equidistributed modulo 1 if for every positive integer p the sequence $\{u_{k+p} - u_k\}$, $k \geq 1$ is equidistributed modulo 1. Equidistribution theorems and uniquely ergodic transformations are intimately connected as pointed out by H Furstenberg. Exploiting these relations it is shown that for any real polynomial $P(x)$ with at least one term of the form ux^n , where u is irrational and $n \geq 1$, the sequence $\{P(k)\}$, $k \geq 1$ is equidistributed modulo 1.

With its well punctuated historical comments and instructive exercises this little but very rich volume offers an enjoyable guided tour of classical harmonic analysis with some scope in trimming its price for the Indian market.

K R Parthasarathy is with the Indian Statistical Institute, New Delhi 110 016.

Basic Biotechnology

S Vijaya



Basic Biotechnology
S Ignacimuthu
Tata McGraw Hill
Publishing Co. Ltd., 1995
pp.317, Rs.135
ISBN 0-07-462162-9

S Ignacimuthu has expertly and compactly packed into one lucid exercise the multifarious topics that we have today come to understand as biotechnology. From the techniques of genetic engineering and its applications, through plant and animal cell culture, agriculture and industrial biotechnology, Ignacimuthu takes his audience into the subjects of gene transfer, healthcare applications and pollution control, all of which have benefited by the advent of biotechnology.

The source of the material is standard and is from well known textbooks written by authorities in their respective fields. In addition the excellent illustrations from these sources go a long way in enhancing the understanding of the reader. The author is to be complimented for very effectively abridging these extensive topics into highly readable material. It is pleasing to see the historical development of every field set down in perspective as he unravels each subject. For example, he describes how the discovery of DNA uptake by bacteria laid the

foundations for the present day sophisticated manipulations practiced by the genetic engineers. The topics are brought up to date with references to very recent developments in each field.

Without doubt, the best chapters are those on plant biotechnology, reflecting the author's familiarity with this field. One cannot but feel on reaching the end of the book that the future of our planet lies in our plants. The unique genetic systems available in plants and the manner in which they have been exploited have been brought out admirably.

The author has painstakingly compiled a valuable list of reading material at the end of each chapter that permits the interested student to pursue whichever chapter captures his or her fancy at greater depth. These include textbooks as well as original articles from journals ranging from *Scientific American* to *Cell*.

My comments are not entirely laudatory. The book does have several errors in the text, a couple of which are cited here: on page 21, it states that to locate particular genes on fragments of a restriction map, Northern blotting can be used; on pages 44 and 47, it states that Southern blot analysis can be used to study expression of genes. Such errors and ambiguities need to be corrected.

The discussion on the ethics of biotechnological applications makes heartening reading. The author rightly

Biotechnology is a term said to have been coined in 1917 by a Hungarian farmer named Ereky. Fermentation by yeasts to produce alcoholic beverages, large scale production of antibiotics and vaccines are some of the familiar examples of application of biotechnology. Biotechnology as conceived initially was essentially an amalgamation of chemical engineering and microbiology. Inputs into the field have always been from observations made in the context of specific questions addressing naturally occurring phenomena; but biotechnology carries these further, to harness practical benefits.

The more recent advances in biotechnology are the results of our understanding of biological processes and their control at the molecular level, an area referred to as molecular biology. One of the important advances in this field which is relevant to biotechnology is the discovery of restriction endonucleases which cleave DNA at specific sequences. Using these enzymes along with a joining or a ligating enzyme, desired pieces of DNA can be brought together and made contiguous. This discovery has enabled the introduction of desired gene sequences into any organism and has led to the creation of a number of *designer* microorganisms in plants and animals. It is worthwhile to remind ourselves that progress in biotechnology is always an outcome of conceptual and technological advances in various disciplines like genetics, biochemistry and several other basic sciences.

The benefits of biotechnology extend from healthcare to pollution management. One example is the remarkable fact that biotechnology has made it possible to produce quantities of growth hormone required for treatment of dwarfism in a child for one year from a couple of litres of bacterial culture (grown for 6-8 hours) instead of the same being obtained from about 75 pituitary glands (got from the same number of human cadavers). In agriculture, genetically modified strains of nitrogen fixing bacteria are being tested for release. Pest control through toxin gene transfer to bacterial strains suitable for infecting plants and frost protection of crops through genetically modified bacteria are some of the other significant developments. It is however necessary to exercise caution while releasing genetically modified microorganisms into the environment. One of the many possible threats to the environment is the production of drug resistant strains of pathogenic organisms.

stresses the importance of preserving human dignity above all other considerations and advocates the practice of biotechnology within the framework of certain value systems.

To conclude, it is my opinion that this excellent book should be added to the list of prescribed textbooks in an essential course

on biotechnology for students of undergraduate programmes in all biology streams.

S Vijaya is with the Centre for Genetic Engineering, Indian Institute of Science, Bangalore 560 012, India.

Acknowledgements

Resonance gratefully acknowledges the help received from the following individuals:

S C Bagchi
C Varughese
Jayant Rao

Annual Subscription Rates

	Personal	Institutional
India	Rs 100	Rs 200
Third World Countries	US \$ 25	US \$ 50
Other Countries	US \$ 50 (Students \$ 25)	US \$ 100

*Send your subscriptions by DD or MO in favour of
"Indian Academy of Sciences" to Circulation Department, Indian Academy of
Sciences, C V Raman Avenue, PB No. 8005, Bangalore 560 080, India.*

*Edited and published by V K Gaur for the Indian Academy of Sciences,
Bangalore 560 080. Printed at Thomson Press (I) Ltd., Faridabad 121 007.*

There's Something in a Name!
Resonance, Vol.1, No.9 (1996)



We featured the life and career of the mathematician Sonya Kovalevskaya in the September issue of *Resonance* (Article in a Box, back cover portrait). While her maiden name was Sofya Krukovskaya she is known in the mathematical world as Sonya Kovalevskaya.

Errata

Resonance, Vol.1, No.5 (1996)

Page 10: The formula for the integral of the Gaussian curvature on a triangle whose sides are geodesics should read thus:

$$\int_{\Delta} (\text{curvature}) = (\text{constant}) \times ((\text{sum of the angles of } \Delta) - \pi)$$

Resonance, Vol.1, No.9 (1996)

Page 2: The text "After Heisenberg's discovery ... in early 1962..." should read "After Heisenberg's discovery ... in early 1926 ..."

Barbara McClintock (16 June 1902 – 2 September 1992) can be described as a visionary in genetics. She received the Nobel prize in 1983 for her work on genetic transposition done in the 1940's.



Barbara McClintock

(1902-1992)